1967

Study of residuals in planned experiments

Francis Gerhard Giesbrecht
Iowa State University

This dissertation has been
microfilmed exactly as received 68-5951

GIESBRECHT, Francis Gerhard, 1935-
STUDY OF RESIDUALS IN PLANNED EXPERIMENTS.

Iowa State University, Ph.D., 1967
Statistics

University Microfilms, Inc., Ann Arbor, Michigan

STUDY OF RESIDUALS IN PLANNED EXPERIMENTS

by

Francis Gerhard Giesbrecht

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1967

TABLE OF CONTENTS

I. INTRODUCTION

A. Statement of the Problem

This dissertation is concerned with the problem of examining the residuals remaining after a conventional least-squares analysis of the data under an assumed linear model has been completed. To be more specific, it deals with the methods of examining the residuals which are defined as follows: Let $\{y_i\}$ be the set of n observed numbers, $\{\alpha_j\}$ a set of p unknown parameters, $\{x_{ij}\}$ a set of np known constants and $\{e_i\}$ a set of n identically distributed $N(0,\sigma^2)$ errors. Suppose that the fitted model

$$y_i = \sum_{j=1} \alpha_j x_{ij} + e_i .$$

The set of residuals $\{z_i\}$ are defined by the n equations

$$z_i = y_i - \sum_{j=1} \hat{\alpha}_j x_{ij} = y_i - Y_i ,$$

where the $\{\hat{\alpha}_i\}$ are those values which minimize the quantity

$$Q = \sum_{i=1}^{n} \{y_i - \sum_{j=1} \hat{\alpha}_j x_{ij}\}^2 .$$

Clearly, the magnitude and structure of the residuals can shed some light on the adequacies of the model, if one can develop objective ways of examining them.

The related problem of finding appropriate transformations of the observed $\{y_i\}$ if the residuals warrant the rejection of the original model is also examined.

Two distinct techniques for doing this will be developed. The first of these depends on the definition of a family of transformations and the use of the principle of maximum likelihood estimation to find the appropriate member of the family. The second procedure selects the monotone transformation which maximizes the correlation between the ordered residuals and the normal order statistics.

B. Review of Literature

The classical least-squares estimates of the $\{\alpha_i\}$ are unquestionably satisfactory under the conditions described in Section A. This set of conditions will be referred to as the ideal statistical conditions, or simply as the ideal conditions. Statisticians have long been interested in the problem of detecting departures from the ideal conditions.

One popular approach to this problem has led to the development of a set of techniques commonly grouped under the heading of "Rejection of Outliers". A historical outline of the progress in this area may be found in recent papers by Anscombe (1960) and Grubbs (1950). The statistical properties of several criteria for rejecting outliers under two relatively distinct mathematical models are discussed in papers by Dixon (1950) and Grubbs (1950). In both models it is assumed that in a sample of n observations, all but a few are drawn from a $N(\mu, \sigma^2)$ population. Under one of the models, the remaining observations are assumed to be from a $N(\mu', \sigma^2)$ population, and under the other, from a $N(\mu, \sigma'^2)$ population. The problem of examining outliers has been dealt with by both of these authors as something in the nature of a test of significance. The treatment in the paper by Anscombe (1960), however, is based on the principle that rules for rejecting outliers are not significance tests. The argument is that the statistician may not be interested in studying whether spurious values occur or not, but rather in guarding himself from their adverse effects. The paper is directed chiefly at the problem of examining the mean square error of estimators following application of a rejection rule. Dixon (1960) demonstrated that satisfactory estimates of the mean could be obtained from normal samples if one or more extreme values were discarded. The cost was a small loss of efficiency.

An alternative to the practice of rejecting outliers is to modify them. Dixon (1960) has investigated the merits of a technique usually credited to C. P. Winsor in which extreme values are replaced with the next largest (or smallest) observation. He found that the cost of this procedure was also only a small loss of efficiency for estimates of the mean when samples were from a normal population.

The technique of modifying extreme values rather than rejecting them outright was extended by Tukey (1962) to the case where the data is in the form of an $r \times c$ array. The extension is not completely straightforward because of difficulties in determining which values must be modified. Basically the approach is to fit a general mean, row effects and column effects, calculate the residuals, order them, plot them against typical values for normal order statistics, draw a straight line through the result and finally modify the observations which correspond to residuals that deviate too much from the line. This procedure is valuable for detecting cases in which the error deviations come from a distribution other than the normal as well as cases in which several of the observed numbers are outliers.

Another approach, which was first developed by Tukey (1949) for the $r \times c$ table is to perform a test for non-additivity. This was extended to more complex classifications by Moore and Tukey (1954) and Tukey (1955). The test is based on a method for removing a special contrast with one degree of freedom from the complete set of error contrasts and comparing it with an error mean square computed from the remaining error contrasts. The test statistic is a function of the quantity $\sum_i z_i Y_i^2$.

Additional techniques for examining the residuals in order to detect various types of deviations from the ideal conditions, based on the third and fourth powers of the residauls are developed in papers by Anscombe (1961) and Anscombe and Tukey (1963). Anscombe (1961) uses the ratio of the sum of the cubes of the residuals to the cube of the estimated standard deviation to develop a test statistic which is analogous to the g_1 statistic discussed by Fisher (1958) for the simple random sample. Anscombe (1961) also uses the ratio of the fourth moment of the residuals to the square of the second moment to generate a statistic analogous to Fisher's g_2 statistic. The difference between the statistics discussed by Anscombe (1961) and those discussed by Fisher (1958) is that the latter apply only to the simple homogeneous sample, while the former are available for more general patterns of observations.

Anscombe (1955) presents a discussion of the statistic $\sum_i z_i^2 Y_i$ as a measure of heteroscedasticity. Some of the properties of this statistic are presented in more detail by Anscombe (1961). This statistic was suggested on the basis of an examination of the plot of the $\{z_i\}$ against the $\{Y_i\}$ on a scatter diagram. Since one of the first steps in any examination of the residuals is to make sure that the sum is zero and since $\sum_i z_i(Y_i) = 0$, it follows that the ordinate (residuals) will have zero mean and zero linear component of regression on the $\{Y_i\}$. If the variance of the original $\{y_i\}$ changes progressively with the mean, the variance of the $\{z_i\}$ will change progressively with the $\{Y_i\}$ and the points may have a wedge-shaped outline, suggesting the statistic. It should also be mentioned that if there appears to be a curvilinear relationship between the $\{z_i\}$ and the $\{Y_i\}$, attention is immediately directed to the statistic $\sum_i z_i Y_i^2$. This leads to Tukey's test for non-additivity.

The problem of finding a suitable transformation of the data if non-additivity is detected is discussed in papers by Tukey (1949), Moore and Tukey (1954), Tukey (1957) and Anscombe (1961). The approach is to find a transformation which will reduce the test statistic for Tukey's test for non-additivity to a satisfactory level. An alternative approach to the problem was presented in a paper by Box and Cox (1964). This approach involves the use of the maximum likelihood principle to choose a particular member of the family of power transformations. A third approach to the problem, which allows a much more general family of transformations has recently been explored in an unpublished paper by Kruskal. The procedure used in this paper is based on finding the monotone transformation which minimizes a squared-residual criterion. The minimization is accomplished by using iterative numerical techniques.

II. MOMENTS OF THE RESIDUALS WHEN SAMPLING FROM AN INFINITE POPULATION

A. Introduction

In this chapter it will be assumed that a set of n numbers, denoted by $\{y_i\}$ have been observed. It will also be assumed that the model

$$y_i = \sum_{j=1}^{p} \alpha_j x_{ij} + e_i \tag{2.1}$$

for i = 1, 2, ..., n is appropriate, where $\{\alpha_j\}$ is a set of p unknown parameters, $\{x_{ij}\}$ a set of np known constants and $\{e_i\}$ a set of n independent and identically distributed errors. The set of residuals $\{z_i\}$ are defined by the n equations

$$z_i = y_i - \sum_{j=1}^{p} \hat{\alpha}_j x_{ij} \tag{2.2}$$

where the $\{\hat{\alpha}_j\}$ are those values which minimize the quantity

$$Q = \sum_{i=1}^{n} \{y_i - \sum_{j=1}^{p} \hat{\alpha}_j x_{ij}\}^2 . \tag{2.3}$$

By setting the p partials of Q with respect to the elements of $\{\hat{\alpha}_j\}$ equal to zero one obtains the equations

$$\sum_{i=1}^{n} \sum_{\ell=1} \hat{\alpha}_\ell x_{i\ell} x_{ij} = \sum_{i=1}^{n} y_i x_{ij} , \tag{2.4}$$

for j = 1, 2, ..., p. Since the $\{x_{ij}\}$ are known constants, this set of linear functions can be solved for the $\{\hat{\alpha}_j\}$ in terms of the $\{y_i\}$. It follows that the n quantities defined by the linear functions

$$\sum_{j=1}^{p} \hat{\alpha}_j \, x_{ij}$$

for i = 1, 2, ..., n are also linear function of the $\{y_i\}$. Consequently the $\{z_i\}$ defined by (2.2) are linear functions of the $\{y_i\}$ and can be written as

$$z_i = \sum_{j=1}^{n} q_{ij} \, y_j \tag{2.5}$$

for i = 1, 2, ..., n.

The purpose of this chapter is to develop a method for obtaining the expected value of expressions like

$$\prod_{a=1}^{m} z_{i_a} = \prod_{a=1}^{m} \left[\sum_{j=1}^{n} q_{i_a j} \, y_j \right]$$

in a compact form in terms of the moments of the $\{y_i\}$. The method to be developed is valid even when some or all of the $\{i_a\}$ are unequal. The formulas are most conveniently given in terms of the cumulants of the $\{y_i\}$ rather than in terms of moments. Formulas to convert cumulants to moments and vice versa can be found in Kendall (1952).

B. Preliminary Definitions and Concepts

In this chapter the expression K_m will be used to denote the m-th cumulant and μ'_m the m-th moment about zero of the $\{y_i\}$. It will be assumed throughout that the distribution of the $\{y_i\}$ is such that all moments mentioned exist.

It is shown in Kendall (1952) that

$$\mu'_m = \sum_{r=1} \sum \left(\frac{K_{p_1}}{p_1!}\right)^{a_1} \left(\frac{K_{p_2}}{p_2!}\right)^{a_2} --- \left(\frac{K_{p_r}}{p_r!}\right)^{a_r} \frac{m!}{a_1!a_2!\ldots a_r!} \qquad (2.6)$$

where the second summation is over all non-negative values of the $\{a_j\}$ such that

$$a_1p_1 + a_2p_2 + \ldots + a_rp_r = m \ . \qquad (2.7)$$

An alternative form is

$$\mu'_m = \sum \left(\frac{K_1}{1!}\right)^{a_1} \left(\frac{K_2}{2!}\right)^{a_2} --- \left(\frac{K_m}{m!}\right)^{a_m} \frac{m!}{a_1!a_2!\ldots a_m!} \ , \qquad (2.8)$$

where the summation extends over all non-negative sets of $\{a_j\}$ such that

$$\sum_{j=1}^{m} ja_j = m \qquad (2.9)$$

The $\{a_j\}$ are readily obtained from the partitions of m . a_j is the number of times the integer j occurs among the parts in a given partition of m . For example,

$$\begin{aligned} \mu'_4 = & \left(\frac{K_1}{1!}\right)^4 \frac{4!}{4!} + \left(\frac{K_1}{1!}\right)^2 \left(\frac{K_2}{2!}\right) \frac{4}{2!1!} \\ & + \left(\frac{K_2}{2!}\right)^2 \frac{4!}{2!} + \left(\frac{K_1}{1!}\right)\left(\frac{K_3}{3!}\right) \frac{4}{1!1!} \\ & + \left(\frac{K_4}{4!}\right) \frac{4!}{1!} \\ = & \ K_1^4 + 6\, K_2K_1^2 + 3K_2^2 + 4K_3K_1 + K_4 \ . \end{aligned}$$

Now recall that if c is a constant and x a random variable then the m-th cumulant of cx is c^m times the m-th cumulant of x . Also if x and y are independent random variables then the m-th cumulant of their sum is equal to the sum of the m-th cumulants of x and y . Consequently, if $K_m(z_i)$ represents the m-th cumulant of z_i and K_m the m-th cumulant of e_i then

$$E[z_i^m] = \mu'_m(z_i)$$

$$= \sum \left(\frac{K_1(z_i)}{1!}\right)^{a_1} \left(\frac{K_2(z_i)}{2!}\right)^{a_2} \cdots \left(\frac{K_m(z_i)}{m!}\right)^{a_m} \frac{m!}{a_1!a_2!\ldots a_m!} \tag{2.10}$$

$$= \sum \left(\frac{\sum_j q_{ij} K_1}{1!}\right)^{a_1} \cdot \left(\frac{\sum_j q_{ij}^2 K_2}{2!}\right)^{a_2} \cdots \left(\frac{\sum_j q_{ij}^m K_m}{m!}\right)^{a_m} \frac{m!}{a_1!a_2!\ldots a_m!} ,$$

where the initial summation in both of these expressions is over all non-negative sets of $\{a_j\}$ subject to the condition that

$$\sum_{j=1}^{m} ja_j = m .$$

This can be written more compactly as

$$E[z_i^m] = \sum \prod_{\ell=1}^{m} \left(\frac{\sum_j^n q_{ij}^\ell K_\ell}{\ell!}\right)^{a_\ell} \frac{m!}{a_1!a_2!\ldots a_m!} \tag{2.11}$$

An example of the use of this expansion is

$$E[z_i^3] = \sum_j q_{ij}^3 K_3 + 3(\sum_j q_{ij}^2)(\sum_j q_{ij}) K_2K_1 + (\sum_j q_{ij})^3 K_1^3 .$$

The purpose of this chapter is to show that there exists a direct extension to the case where the i subscripts may be unequal. If one defines

$$s_0 = 0$$

$$s_1 = a_1$$

$$s_2 = a_1 + 2a_2$$

$$= s_1 + 2a_2$$

$$\vdots$$

$$s_m = s_{m-1} + ma_m$$

then this extension will be shown to be

$$E[\prod_{r=1}^{m} z_{i_r}] = \sum_1 \sum_2 \prod_{r=1}^{m} \left\{ \left[\sum_j \left(\prod_{h=s_{r-1}+1}^{s_{r-1}+r} q_{\ell_h j} \right) K_r \right] \cdots \right.$$

$$\left. \left[\sum_j \left(\prod_{h=s_r-r+1}^{s_r} q_{\ell_h j} \right) K_r \right] \right\} \tag{2.12}$$

$$= \sum_1 \sum_2 \prod_{r=1}^{m} \left\{ \left[\sum_j \left(\prod_{h=s_{r-1}+1}^{s_{r-1}+r} q_{\ell_h j} \right) \right] \cdots \right.$$

$$\left. \left[\sum_j \left(\prod_{h=s_r-r+1}^{s_r} q_{\ell_h j} \right) \right] K_r^{a_r} \right\}$$

where $\sum_1$ is over all non-negative integral $\{a_r\}$ subject to the condition that

$$\sum_{r=1}^{m} ra_r = m$$

and $\sum_2$ is over all possible distinct assignments, denoted by the $\{\ell_h\}$, of the $\{i_r\}$ subscripts to $\sum_{r=1}^{m} a_r$ groups in which a_1 contain one element, a_2 contain two elements, ... and a_m contain m elements. The permutations of groups of ℓ's and permutations of ℓ's within groups of ℓ's are not distinguished.

The meaning of this can be clarified by examining $E[z_i^2\ z_{i'}]$ and $E[z_i\ z_{i'}\ z_{i''}]$ where i, i' and i" are all unequal and comparing with the value for $E[z_i^3]$ given previously.

$$\begin{aligned} E[z_i\ z_{i'}\ z_{i''}] &= \sum_j q_{ij}\ q_{i'j}\ q_{i''j}\ K_3 + \\ &\left\{\left(\sum_j q_{ij}\ q_{i'j}\right)\left(\sum_j q_{i''j}\right) + \left(\sum_j q_{ij}\ q_{i''j}\right)\left(\sum_j q_{i'j}\right)\right. \\ &\left. + \left(\sum_j q_{i'j}\ q_{i''j}\right)\left(\sum_j q_{ij}\right)\right\} K_2K_1 \\ &+ \left(\sum_j q_{ij}\right)\left(\sum_j q_{i'j}\right)\left(\sum_j q_{i''j}\right) K_1^3 \end{aligned}$$

and

$$E[z_i^2 z_{i'}] = \sum_j q_{ij}^2 q_{i'j} K_3 + \left\{\left(\sum_j q_{ij}^2\right)\left(\sum_j q_{i'j}\right)\right.$$

$$\left. + 2\left(\sum_j q_{ij} q_{i'j}\right)\left(\sum_j q_{ij}\right)\right\} K_2 K_1 + \left(\sum_j q_{ij}\right)^2 \left(\sum_j q_{i'j}\right) K_1^3 .$$

C. Proof of General Expansion

The purpose of the present section is to establish the validity of the expression

$$E\left[\prod_{r=1}^{m} z_{i_r}\right] = \sum\nolimits_1 \sum\nolimits_2 \prod_{r=1}^{m} \left\{\left[\sum_j \left(\prod_{h=s_{r-1}+1}^{s_{r-1}+r} q_{\ell_h j}\right)\right] \cdots\right.$$

$$\left.\left[\sum_j \left(\prod_{h=s_r-r+1}^{s_r} q_{\ell_h j}\right)\right] K_r^{a_r}\right\} , \tag{2.13}$$

where the sum $\sum_1$ is over all non-negative integral $\{a_r\}$ subject to the condition that

$$\sum_{r=1}^{m} r a_r = m$$

and the sum $\sum_2$ is over all possible distinct assignments, denoted by the $\{\ell_h\}$, of the i_r subscripts of the $\sum_{r=1}^{m} a_r$ groups in which a_1 contain one element each, a_2 contain two elements each, ... and a_m contain m elements. The $\{s_r\}$ are defined by the series of equations

$$s_r = \sum_{i=1}^{r} i a_i \quad \text{with} \quad s_0 = 0 .$$

The permutation of groups of ℓ's and permutations within groups of ℓ's are not distinguished. There is no restriction on the nature $\{i_r\}$, i.e., they may or may not all be distinct.

The procedure used is to obtain the appropriate term, i.e., the coefficient of $\prod_{r=1}^{m} t_r$ in the joint moment generating function of the $\{z_{i_r}\}$. This function is defined to be

$$M(z_{i_1} \ldots z_{i_m}) = E\left[e^{\sum_{r=1}^{m} t_r z_{i_r}}\right] . \tag{2.14}$$

However $z_{i_r} = \sum_j q_{i_r j} e_j$. Therefore,

$$\sum_{r=1}^{m} t_r z_{i_r} = \sum_{r=1}^{m} t_r (\sum_j q_{i_r j}) e_j$$

$$= \sum_j (\sum_{r=1}^{m} t_r q_{i_r j}) e_j$$

$$= \sum_j v_j e_j .$$

Hence

$$M(z_{i_1} \ldots z_{i_m}) = E\left[e^{\sum_j v_j e_j}\right] .$$

However, the moment generating function of e_j is defined as

$$E[e^{ve_j}] = 1 + \mu_1' \, v + \mu_2' \, \frac{v^2}{2!} + \mu_3' \, \frac{v^3}{3!} + \ldots$$

$$= e^{K_1 v + K_2 \frac{v^2}{2!} + K_3 \frac{v^3}{3!} \ldots} \quad ,$$

where μ_r' and K_r are the n'th moment and cumulant, respectively, of e_j .

For independent $\{e_j\}$, it follows immediately that

$$M(z_{i_1} \ldots z_{i_m}) = e^{(\Sigma v_j) K_1 + \frac{1}{2!}(\Sigma v_j^2) K_2 + \frac{1}{3!}(\Sigma v_j^3) K_3 \ldots} \tag{2.15}$$

$$= 1 + \{(\sum v_j) K_1 + \frac{1}{2!}(\sum v_j^2) K_2 + \ldots\}$$

$$+ \frac{1}{2!} \{(\sum v_j) K_1 + \frac{1}{2!}(\sum v_j^2) K_2 + \ldots\}^2$$

$$+ \frac{1}{3!} \{(\sum v_j) K_1 + \frac{1}{2!}(\sum v_j^2) K_2 + \ldots\}^3$$

$$+ \ldots$$

$$= 1 + \{ \sum_{h=1}^{\infty} \frac{1}{h!} \sum_{j=1}^{n} (\sum_{r=1}^{m} t_r \, q_{i_r j})^h K_h\}$$

$$+ \frac{1}{2!} \{ \sum_{h=1}^{\infty} \frac{1}{h!} \sum_{j=1}^{n} (\sum_{r=1}^{m} t_r \, q_{i_r j})^h K_h\}^2$$

$$+ \frac{1}{3!} \{ \sum_{h=1}^{\infty} \frac{1}{h!} \sum_{j=1}^{n} (\sum_{r=1}^{m} t_r \, q_{i_r j})^h K_h\}^3$$

$$+ \ldots \; ..$$

Under the usual condition that $\sum_j q_{ij} = 0$ (the residuals sum to zero), the coefficient of K_1 is zero in the above expression. The result is that the above expression can be written with the subscript h starting from 2 rather than 1. The expected value of the product $\prod_{r=1}^{m} z_{i_r}$ is obtained by picking out the coefficients of $\prod_{r=1}^{m} t_r$. On expanding the above expression the term $\prod_{r=1}^{m} t_r$ will be seen to occur among the terms derived from each of the first m lines. The coefficient of $\Pi\, t_r$ in the series of terms from the first line is

$$\frac{1}{m!}\left[\sum_{j=1}^{n} \frac{m!}{1!1!\ldots1!} q_{i_1 j} \cdots q_{i_m j}\right] .$$

The coefficient of $\Pi\, t_r$ in the series of terms from the second line is

$$\frac{2!}{2!}\left\{\frac{1}{2!}\frac{1}{(m-2)!}\sum\nolimits_2\left[\sum_{j_1=1}^{n}\frac{2!}{1!1!} q_{\ell_1 j_1} q_{\ell_2 j_1}\right]\right.$$

$$\left.\left[\sum_{j_2=1}^{n}\frac{(m-2)!}{1!\ldots1!} q_{\ell_3 j_2} \cdots q_{\ell_m j_2}\right]\right\} K_2 K_{m-2}$$

$$= \sum\nolimits_2\left[\sum_{j_1=1}^{n} q_{\ell_1 j_1} q_{\ell_2 j_1}\right]\left[\sum_{j_2=1}^{n} q_{\ell_3 j_2} \cdots q_{\ell_m j_2}\right] K_2 K_{m-2} ,$$

where $\sum_2$ is over the possible associations of the $\{i_r\}$ subscripts and the $\{\ell_h\}$ subscripts. Similarly the coefficient of $\Pi\, t_r$ in the series of terms from the third line is

$$\sum_2 \Big[\sum_{j_1=1}^{n} q_{\ell_1 j_1} q_{\ell_2 j_1} q_{\ell_3 j_1} \Big] \Big[\sum_{j_2=1}^{n} q_{\ell_4 j_2} \cdots q_{\ell_m j_2} \Big] K_3 K_{m-3} \ .$$

The series of terms from the fourth line will yield two expressions, one containing $K_4 K_{m-4}$ and the other containing $K_2 K_2 K_{m-4}$.

However, the sequence of terms obtained in this process is identical to that given on the right hand side of (2.13). Consequently, the desired relationship is established. Since the $\{i_r\}$ subscripts were not restricted, it follows that the results are also valid for all cases where some of these values may be equal. In fact, in the special case where all $\{i_r\}$ subscripts are equal was established earlier by a simpler argument.

It has been assumed that $\sum_j q_{ij} = 0$ for all values of i . The modifications of the results to allow more general $\{q_{ij}\}$ are obvious and will not be given.

D. Moments of the Residuals

The formula derived in the previous section yields the expected value of all powers and products of the residuals in terms of cumulants. For terms of order ten or less, these can be converted to expansions in terms of moments by using formulas given in Kendall (1952). If higher order terms are required, the general conversion formula must be used.

The general formula derived above will now be used to obtain some of the moments of the residuals needed most frequently. It will be assumed that the original model was such that the residuals always sum to zero, i.e., $\sum_j q_{ij} = 0$.

$$E[z_i^2] = \sum_j q_{ij}^2 \; K_2$$

and more generally

$$E[z_i z_{i'}] = \sum_j q_{ij} \; q_{i'j} \; K_2 \; .$$

$$E[z_i^3] = \sum_j q_{ij}^3 \; K_3 \; ,$$

$$E[z_i^2 \; z_{i'}] = \sum_j q_{ij}^2 \; q_{i'j} \; K_3$$

and

$$E[z_i \; z_{i'} \; z_{i''}] = \sum_j q_{ij} \; q_{i'j} \; q_{i''j} \; K_3 \; .$$

$$E[z_i^4] = \sum_j q_{ij}^4 \; K_4 + 3(\sum_j q_{ij}^2)^2 \; K_2^2 \; ,$$

$$E[z_i^3 \; z_{i'}] = \sum_j q_{ij}^3 \; q_{i'j} \; K_4 + 3(\sum_j q_{ij}^2)(\sum_j q_{ij} \; q_{i'j}) \; K_2^2 \; ,$$

$$E[z_i^2 \; z_{i'}^2] = \sum_j q_{ij}^2 \; q_{i'j}^2 \; K_4 + [(\sum_j q_{ij}^2)(\sum_j q_{ij}^2) + 2(\sum_j q_{ij} \; q_{i'j})^2] \; K_2^2 \; ,$$

$$E]z_i^2 \; z_{i'} \; z_{i''}] = \sum_j q_{ij}^2 \; q_{i'j} \; q_{i''j} \; K_4 + [(\sum_j q_{ij}^2)(\sum_j q_{i'j} \; q_{i''j})$$

$$+ 2(\sum_j q_{ij} \; q_{i'j})(\sum_j q_{ij} \; q_{i''j})] \; K_2^2$$

and

$$
\begin{aligned}
E[z_{i_1}\ z_{i_2}\ z_{i_3}\ z_{i_4}] = &\sum_j q_{i_1j}\ q_{i_2j}\ q_{i_3j}\ q_{i_4j}\ K_4 \\
&+ [(\sum_j q_{i_1j}\ q_{i_2j})(\sum_j q_{i_3j}\ q_{i_4j}) \\
&+ (\sum_j q_{i_1j}\ q_{i_3j})(\sum_j q_{i_2j}\ q_{i_4j}) \\
&+ (\sum_j q_{i_1j}\ q_{i_4j})(\sum_j q_{i_2j}\ q_{i_3j})]\ K_2^2\ .
\end{aligned}
$$

III. A MEASURE OF SKEWNESS WHEN A LINEAR MODEL HOLDS

A. Introduction

The statistic

$$g_1 = \frac{\sum_i z_i^3}{s^3 \sum_{ij} q_{ij}^3} \tag{3.1}$$

where

$$s^2 = \frac{\sum_i z_i^2}{\sum_{ij} q_{ij}^2}$$

and

$$\sum_{ij}^{n} q_{ij}^3 \neq 0$$

has been proposed by Anscombe (1961) as a suitable criterion to test whether the errors in the original observations are from a skewed distribution. In the same paper it is shown that if the errors are NID(0, σ^2) then:

1. The mean or expected value of g_1 is zero.
2. The exact variance, equal to the expected value of $(g_1)^2$ is

$$\frac{[6 \sum_{ij} q_{ij}^3 + 9 \sum_{ij} q_{ij}\, q_{ii}\, q_{jj}]\, \nu^2}{[\sum_{ij} q_{ij}^3]^2\, (\nu+2)(\nu+4)},$$

where

$$\nu = \sum_{ij} q_{ij}^2 .$$

3. The expected value of $(g_1)^3$ is zero

4. The exact expression for the expected value of $(g_1)^4$ is

$$\frac{108\ \nu^5}{(\nu+2)(\nu+4)(\nu+6)(\nu+8)(\nu+10)\,[\sum_{ij} q_{ij}^3]^4}\Bigg\{\Big[\sum_{ij} q_{ij}^3\Big]^2$$

$$+ 18 \sum_{i,j,k,h} q_{ij}^2\, q_{kk}^2\, q_{ik}\, q_{jh} + 12 \sum_{i,j,k,h} q_{ij}\, q_{ik}\, q_{ih}\, q_{jk}\, q_{jh}\, q_{kh}$$

$$+ 36 \sum_{i,j,k,h} q_{ij}\, q_{jk}\, q_{jh}\, q_{kk}^2\, q_{ii} + 18 \sum_{i,j,k,h} q_{ik}\, q_{jh}\, q_{kk}^2\, q_{ii}\, q_{jj}$$

$$+ 6 \sum_{i,j,k,h} q_{ih}\, q_{jh}\, q_{kh}\, q_{ii}\, q_{jj}\, q_{kk} + 3 \sum_{ij} q_{ij}\, q_{ii}\, q_{jj} \sum_{kh} q_{kh}^3$$

$$+ \frac{9}{4}\Big(\sum_{ij} q_{ij}\, q_{ii}\, q_{jj}\Big)^2\Bigg\} \ .$$

The object of the present chapter is to examine the behavior of g_1 when the assumption of normally distributed errors is relaxed.

The assumptions are that an additive model holds and that the errors are identically distributed, independent random variables. It will also be assumed that the error distribution is such that the required moments exist. The method of approach will be to replace g_1 by a series approximation and then examine the properties of this series. Moments of this series can then be evaluated, using the methods developed in Chapter II.

B. Theoretical Development

The derivation in this section will be restricted to the case where the sampling scheme or experimental design is such that $\sum_{i}^{n} q_{ij} = 0$ for all j and q_{ii} is constant for all i. The first condition is true for all classification models, and is a consequence of the fact that the parameter set can be chosen so that one parameter is a general mean and enters with coefficient one. It implies that the residuals sum to zero. If this restriction is not satisfied, a first step in the examination of the residuals could well be to consider the statistic $(\Sigma z_i)^2/(\Sigma z_i{}^2)$. Properties of this statistic are not considered in this thesis though the same reasoning could be applied. The second restriction implies that the residuals have equal variances, under the assumption of homogeneous errors. Consequently one can write

$$E \sum_{i}^{n} z_i^2 = \nu\mu_2 ,$$

where $\nu = \sum_{ij} q_{ij}^2$.

Let

$$\delta = \sum_{i} z_i^2 - \nu\mu_2 .$$

Then if $\sum_{ij}^{n} q_{ij}^3 \neq 0$, one can write

$$g_1 = \frac{(\sum_{ij} q_{ij}^2)^{3/2} (\sum_{i} z_i^3)}{(\sum_{ij} q_{ij}^3)(\nu\mu_2)^{3/2}} [1 + \frac{\delta}{\nu\mu_2}]^{-3/2} . \tag{3.2}$$

Now recall that $[1 + \frac{\delta}{\nu\mu_2}]^{-x}$ can be expanded into a converging power series if $|\frac{\delta}{\nu\mu_2}| < 1$. Since the variance of δ is equal to

$$E[\delta^2] = \frac{\nu^2}{n} K_4 + 2\nu(\mu_2)^2 ,$$

where K_4 is the fourth cumulant of the distribution of the errors, it follows that

$$E\left[\left(\frac{\delta}{\nu\mu_2}\right)^2\right] = \frac{K_4}{n\mu_2^2} + \frac{2}{\nu} .$$

One can surmise therefore, if the number of degrees of freedom for error is at all appreciable e.g. greater than 6, say, that the probability of $|\frac{\delta}{\nu\mu_2}| \geqslant 1$ is small, so that the expansion is a reasonable approximation. There are, however, non-trivial theoretical obscurities about such approximations which are widely used in statistical reasoning.

It follows that an approximation for g_1 obtained by expanding $[1 + \frac{\delta}{\nu\mu_2}]^{-3/2}$ and retaining the first four terms, is

$$g_1^* = \frac{(\sum_{ij} q_{ij}^2)^{3/2} \, (\sum_i z_i^3)}{(\sum_{ij} q_{ij}^3)(\nu\mu_2)^{3/2}} \left[1 - \frac{3}{2}\,\frac{\delta}{\nu\mu_2} + \frac{15}{8}\,\frac{\delta^2}{(\nu\mu_2)^2} - \frac{35}{16}\,\frac{\delta^3}{(\nu\mu_2)^3}\right] .$$

More accurate approximations can be obtained by retaining more terms in the series expansion. The expected value of g_1^* is equal to

$$\left.\frac{(\sum_{ij} q_{ij}^2)^{3/2}}{(\sum_{ij} q_{ij}^3)(\nu\mu_2)^{3/2}}\right\} E[\sum_i z_i^3] \tag{3.4}$$

$$- \left(\frac{3}{2}\right) \frac{E[(\Sigma z_i^3)(\Sigma z_i^2)] - \nu\mu_2 E[\Sigma z_i^3]}{\nu\mu_2}$$

$$+ \left(\frac{15}{8}\right) \frac{E[(\Sigma z_i^3)(\Sigma z_i^2)(\Sigma z_i^2)] - 2(\nu\mu_2) E[(\Sigma z_i^3)(\Sigma z_i^2)] + (\nu\mu_2)^2 E[\Sigma z_i^3]}{(\nu\mu_2)^2}$$

$$- \left(\frac{35}{16}\right) \frac{\left[E[(\Sigma z_i^3)(\Sigma z_i^2)(\Sigma z_i^2)(\Sigma z_i^2)] - 3(\nu\mu_2) E[(\Sigma z_i^3)(\Sigma z_i^2)(\Sigma z_i^2)]\right.}{(\nu\mu_2)^3}$$

$$\left.\underline{\left. + 3(\nu\mu_2)^2 E[(\Sigma z_i^3)(\Sigma z_i^2)] - (\nu\mu_2)^3 E[\Sigma z_i^3]\right]}\right\} .$$

The necessary product moments of the residuals can be evaluated by using the method developed in Chapter II. Since the expected values of the product moments of the residuals will involve special functions of the $\{q_{ij}\}$ a special system of notation will be used for these terms. This will consist of symbolic expressions like (112)(1233) used to represent the function

$$\sum_{i_1} \sum_{i_2} \sum_{i_3} (\sum_{j_1} q_{i_1j_1}^2 q_{i_2j_1})(\sum_{j_2} q_{i_1j_2} q_{i_2j_2} q_{i_3j_2}^2) .$$

Similarly (111)(22) will be used to denote the function

$$\sum_{i_1} \sum_{i_2} (\sum_{j_1} q^3_{i_1 j_1})(\sum_{j_2} q^2_{i_2 j_2}) .$$

The association immediately becomes obvious if one notices the correspondence between the numbers in the symbolic expressions and the numerical subscripts on the first or i subscripts on the $\{q_{ij}\}$ elements. The parentheses are used to specify the second or j subscript on the $\{q_{ij}\}$ elements.

From the results in Chapter II it follows immediately that

$$E[\sum z_i^3] = \sum_i E[z_i^3] \tag{3.5}$$

$$= \sum_i [\sum_j q^3_{ij} K_3 + 3(\sum_{j_1} q^2_{ij_1})(\sum_{j_2} q_{ij_2}) K_2K_1$$

$$+ (\sum_{j_1} q_{ij_1})(\sum_{j_2} q_{ij_2})(\sum_{j_3} q_{ij_3}) K_1^3$$

$$= \sum_{ij} q^3_{ij} K_3$$

$$= (111) K_3 .$$

Any term involving K_1 can be dropped because it is always associated with the term $(\sum_j q_{ij})$ which was assumed to be zero. The next term required by (3.4) is the expected value of $(\sum z_i^3)(\sum z_i^2)$.

$$E[\sum_{i_1} z_{i_1}^3)(\sum_{i_2} z_{i_2}^2)] = \sum_{i_1}\sum_{i_2} E[z_{i_1}^3 \; z_{i_2}^2] \tag{3.6}$$

$$= (11122)\ K_5 + [(111)(22) + 6(112)(12)\ K_3K_2 \ .$$

The term (122)(11) in the coefficient on K_3K_2 was discarded since it denotes the sum

$$\sum_{i_1}\sum_{i_2}\sum_{j_1}(q_{i_1j_1}\ q_{i_2j_1}^2)(\sum_{j_2} q_{i_1j_2}^2) \ ,$$

which is zero under the restrictions put onto the $\{q_{ij}\}$ elements. The term (111)(22) can only occur once since there is only one assignment of the 3 i_1 subscripts and 2 i_2 subscripts to 2 sets with the first containing 3 i_1's and the second 2 i_2's. Similarly (112)(12) has the coefficient 6 since 3 i_1's and 2 i_2's can be assigned to a set containing 2 i_1's and one i_2 and another set containing one i_1 and one i_2 in 6 ways. The coefficient on (122)(11) would be 3 since the odd i_1 can be selected in 3 ways. A check to see that all terms have been accounted for is obtained by adding the numerical coefficients. The coefficient on (111)(22) is one, on (112)(12) is 6 and on (122)(11) is 3. The sum 1 + 6 + 3 = 10 is the number of ways 5 objects can be assigned to groups of 3 and 2, i.e., $\frac{5!}{3!2!}$.

$$E[(\sum_{i_1} z_{i_1}^3)(\sum_{i_2} z_{i_2}^2)(\sum_{i_3} z_{i_3}^2)] \tag{3.7}$$

$$\begin{aligned} &= (1112233)\ K_7 + [2(11122)(33) + 4(11123)(23) + 12(11233)(12)]\ K_5K_2 \\ &\quad + [(111)(2233) + 12(112)(1233) + 12(123)(1123)]\ K_3K_4 \\ &\quad + [(111)(22)(33) + 2(111)(23)(23) + 12(112)(12)(33) \\ &\quad + 24(112)(13)(23) + 24(123)(12)(13)]\ K_3K_2^2\ . \end{aligned}$$

The term (11122)(33) appears with the coefficient 2 in the above expressions because the 2 terms (11122)(33) and (11133)(22) are numerically equal and differ only in that the subscripts on the $\{q_{ij}\}$ elements are permuted. Other equivalent expressions have been combined in the same manner.

Note that the complete coefficient on K_5K_2 is

[2(11122)(33) + 4(11123)(23) + 12(11233)(12) + 3(12233)(11)].

The last term can be deleted since it represents a sum which is equal to zero. As a check, one notes that $2 + 4 + 12 + 3 = 21 = \frac{7!}{5!2!}$.
Similarly the complete coefficient on K_3K_4 is

[(111)(2233) + 12(112)(1233) + 12(123)(1123) + 6(122)(1133)
+ 4(223)(1113)]

and $1 + 12 + 12 + 6 + 4 = 35 = \frac{7!}{4!3!}$. A similar sum is available in each case and was computed as a check to ensure that no non-zero terms were neglected.

$$E[(\sum_{i_1} z_{i_1}^3)(\sum_{i_2} z_{i_2}^2)(\sum_{i_3} z_{i_3}^2)(\sum_{i_4} z_{i_4}^2)] = (111223344)\ K_9 \quad (3.8)$$

$$+ [3(1112233)(44) + 12(1112234)(34) + 18(1122334)(14)]\ K_7K_2$$

$$+ [(111)(223344) + 18(112)(123344) + 36(123)(112344)$$

$$+ 8(234)(111234)]\ K_6K_3$$

$$+ [3(11122)(3344) + 12(11123)(2344) + 24(11234)(1234)$$

$$+ 36(11223)(1344) + 36(12234)(1134)]\ K_5K_4$$

$$+ [72(12)(13)(12344) + 72(12)(23)(11344) + 36(12)(33)(11244)$$

$$+ 72(12)(34)(11234) + 3(22)(33)(11144) + 12(22)(34)(11134)$$

$$+ 6(23)(23)(11144) + 24(23)(34)(11124)]\ K_5K_2^2$$

$$+ [3(111)(22)(3344) + 12(111)(23)(2344) + 18(112)(12)(3344)$$

$$+ 72(112)(13)(2344) + 72(112)(23)(1344) + 36(112)(33)(1244)$$

$$+ 72(112)(34)(1234) + 144(123)(12)(1344) + 144(123)(14)(1234)$$

$$+ 144(123)(24)(1134) + 36(123)(44)(1123) + \cancel{72}(234)(12)(1134)]\ K_4K_3K_2$$

$$+ [4(111)(234)(234) + 72(112)(134)(234) + 24(123)(124)(134)]\ K_3^3$$

$$+ [(111)(22)(33)(44) + 6(111)(23)(23)(44) + 8(111)(23)(24)(34)$$

$$+ 18(112)(12)(33)(44) + 36(112)(13)(34)(34) + 72(112)(13)(23)(44)$$

$$+ 144(112)(13)(24)(34) + 72(123)(12)(13)(44) + 288(123)(12)(14)(34)$$

$$+ 48(234)(12)(13)(14)]\ K_3K_2^3 .$$

Some of the expressions given above can be shortened by applying the fact that the $\{q_{ij}\}$ were restricted by two side conditions. However, if this is done, much of the symmetry in the expressions is lost. By letting $\nu = q_{ii}/n$ for all i and applying the restriction

$\sum_i q_{ij} = 0$ for all j, (3.6) can be rewritten as

$$E(\sum z_{i_1}^3)(\sum z_{i_2}^2)$$

$$= \frac{\nu}{n}\,(111)\;K_5 + (\nu+6)(111)\;K_3K_2\;. \tag{3.9}$$

(3.7) can be rewritten as

$$E[(\sum z_{i_1}^3)(\sum z_{i_2}^2)(\sum z_{i_3}^2) \tag{3.10}$$

$$= \left(\frac{\nu}{n}\right)^2 (111)\;K_7 + [2\left(\frac{\nu^2}{n}\right)\;(111) + 16\left(\frac{\nu}{n}\right)\;(111)]\;K_5K_2$$

$$+ [\left(\frac{\nu^2}{n}\right)\;(111) + 12\left(\frac{\nu}{n}\right)\;(111) + 12(123)(1123)]\;K_3K_4$$

$$+ [\nu^2(111) + 14\nu(111) + 48(111)]\;K_3K_2^2\;.$$

(3.8) can be rewritten as

$$E[(\sum z_{i_1}^3)(\sum z_{i_2}^2)(\sum z_{i_3}^2)(\sum z_{i_4}^2)] \tag{3.11}$$

$$= \left(\frac{\nu}{n}\right)^3 (111)\;K_9 + (3\nu+30)\;\left(\frac{\nu}{n}\right)^2\;(111)\;K_7K_2$$

$$+ [(\nu+18)\;\left(\frac{\nu}{n}\right)^2\;(111) + 36\left(\frac{\nu}{n}\right)\;(123)(1123) + 8(111222)]\;K_6K_3$$

$$+ [(3\nu+48)\;\left(\frac{\nu}{n}\right)^2\;(111) + 24(11234)(1234) + 36\left(\frac{\nu}{n}\right)\;(123)(1123)]\;K_5K_4$$

$$+ [240\left(\frac{\nu}{n}\right)(111) + 54\nu\left(\frac{\nu}{n}\right)\;(111) + 3\nu^2\left(\frac{\nu}{n}\right)\;(111)]\;K_5K_2^2$$

$$+ [360\left(\frac{\nu}{n}\right) (111) + 54\nu\left(\frac{\nu}{n}\right) (111) + 3\frac{\nu}{n}^3 (111) + 12\frac{\nu}{n}^2 (111)$$

$$+ (360 + 36\nu)(123)(1123)]\ K_4K_3K_2$$

$$+ [4(111)(222) + 72(112)(134)(234) + 24(123)(124)(134)]\ K_3^3$$

$$+ [\nu^3(111) + 24\nu^2(111) + 188\nu(111) + 480(111)]\ K_3K_2^3\ .$$

These results can be summarized by the equation

$$E[g_1^*] = \frac{1}{(111)\mu_2^{3/2}} \left\{(111)\ K_3\right. \tag{3.12}$$

$$- \left(\frac{3}{2}\right) [\left(\frac{\nu}{n}\right) (111)\ K_5 + (\nu{+}6)(111)\ K_3K_2 - \nu(111)\ K_3K_2][\nu\ K_2]^{-1}$$

$$+ \left(\frac{15}{8}\right) [\left(\frac{\nu}{n}\right)^2 (111)\ K_7 + 16\left(\frac{\nu}{n}\right) (111)\ K_5\ K_2$$

$$+ \{\left(\frac{\nu}{n}\right)^2 (111) + 12\left(\frac{\nu}{n}\right) (111) + 12(123)(1123)\}\ K_4K_3$$

$$+ \{2\nu(111) + 48(111)\}\ K_3K_2^2][\nu\ K_2]^{-2}$$

$$- \left(\frac{35}{16}\right) [\left(\frac{\nu}{n}\right)^3 (111)\ K_9 + 30\left(\frac{\nu}{n}\right)^2 (111)\ K_7K_2$$

$$+ \{\nu\left(\frac{\nu}{n}\right)^2 (111) + 18\left(\frac{\nu}{n}\right)^2 (111) + 36\left(\frac{\nu}{n}\right) (123)(1123) + 8(111222)\}\ K_6K_3$$

$$+ \{3\nu\left(\frac{\nu}{n}\right)^2 (111) + 48\left(\frac{\nu}{n}\right)^2 (111) + 36\left(\frac{\nu}{n}\right) (123)(1123)$$

$$+ 24(11234)(1234)\}\ K_5K_4$$

$$+ \{240\left(\frac{\nu}{n}\right)\ (111) + 6\nu\left(\frac{\nu}{n}\right)\ (111)\}\ K_5K_2^2$$

$$+ \{360\left(\frac{\nu}{n}\right)\ (111) + 30\frac{\nu^2}{n}\ (111) + 360(123)(1123)\}\ K_4K_3K_2$$

$$+ \{4(111)(222) + 72(112)(134)(234) + 24(123)(124)(134)\}\ K_3^3$$

$$\left. + \ \{24\nu(111) + 480(111)\}\ K_3K_2^3][\nu\ K_2]^{-3} \right\} .$$

C. Application to Specific Designs

In order to apply these results to a specific design, the 6 functions (111), (123)(1123), (111222), (1234)(11234), (112)(134)(234) and (123)(124)(134) must be evaluated. These are given in Table 1 for the simple random sample of size n and for the b × t classification in Table 2.

An examination of (3.12) in conjunction with Table 1 indicates that for a simple sample of n the δ term leads to a contribution of order zero in n, the δ^2 term leads to a contribution of order minus one in n and the δ^3 term leads to a contribution of order minus 2 in n.

The technique used to evaluate these functions will be illustrated for the function (123)(124)(134) for the simple random sample. In this case we know that

$$q_{ij} = -\frac{1}{n} \text{ for all } i \neq j \text{ and}$$

$$q_{ii} = \frac{n-1}{n} \text{ for all } i.$$

Table 1. Values of selected symmetric function of the $\{q_{ij}\}$ for the simple random sample of size n.

Function	Value
(111)	$\left(\frac{1}{n}\right)^2 (n-1)[(n-1)^2 - 1]$
(123)(1123)	$\left(\frac{1}{n}\right)^4 (n-1)[(n-1)^2 - 1]^2$
(111222)	$\left(\frac{1}{n}\right)^5 (n-1)^2[(n-1)^2 - 1]^2$
(1234)(11234)	$\left(\frac{1}{n}\right)^4 (n-1)[(n-1)^4 - (n-1)^3 + (n-1) -1]$
(112)(134)(234)	$\left(\frac{1}{n}\right)^2 (n-1)(n-2)^2$
(123)(124)(134)	$\left(\frac{1}{n}\right)^3 (n-1)(n-3)[(n-1)^2 - 1]$

Table 2. Values of selected symmetric functions of the $\{q_{ij}\}$ for the b × t cross classification.

Function	Value
(111)	$\left(\frac{1}{b}\right)^2 (b-1)[(b-1)^2 - 1]\left(\frac{1}{t}\right)^2 (t-1)[(t-1)^2 - 1]$
(123)(1123)	$\left(\frac{1}{b}\right)^4 (b-1)[(b-1)^2 - 1]^2\left(\frac{1}{t}\right)^4 (t-1)[(t-1)^2 - 1]^2$
(111222)	$\left(\frac{1}{b}\right)^5 (b-1)^2[(b-1)^2 - 1]^2\left(\frac{1}{t}\right)^5 (t-1)^2[(t-1)^2 - 1]^2$
(1234)(11234)	$\left(\frac{1}{b}\right)^4 (b-1)[(b-1)^4 - (b-1)^3 + (b-1) - 1] \times$
	$\left(\frac{1}{t}\right)^4 (t-1)[(t-1)^4 - (t-1)^3 + (t-1) - 1]$
(112)(134)(234)	$\left(\frac{1}{b}\right)^2 (b-1)(b-2)^2\left(\frac{1}{t}\right)^2 (t-1)(t-2)^2$
(123)(124)(134)	$\left(\frac{1}{b}\right)^3 (b-1)(b-3)[(b-1)^2 - 1] \times$
	$\left(\frac{1}{t}\right)^3 (t-1)(t-3)](t-1)^2 - 1]$

It follows from the definition that

$$(123)(124)(134) = \sum_{i_1} \sum_{i_2} \sum_{i_3} \sum_{i_4} (\sum_{j_1} q_{i_1 j_1} q_{i_2 j_1} q_{i_3 j_1})(\sum_{j_2} q_{i_1 j_2} q_{i_2 j_2} q_{i_4 j_2})$$

$$(\sum_{j_3} q_{i_1 j_3} q_{i_3 j_3} q_{i_4 j_3}) \tag{3.13}$$

By interchanging the order of summations and using the fact that $\sum_j q_{ij} q_{kj} = q_{ki}$, this can be reduced to

$$\sum_{i_1} \sum_{j_1} \sum_{j_2} \sum_{j_3} q_{i_1 j_1} q_{i_1 j_2} q_{i_1 j_3} q_{j_1 j_2} q_{j_1 j_3} q_{j_2 j_3} .$$

This in turn can be written as

$$\sum_{i_1 \neq j_1} \sum_{j_2} \sum_{j_3} q_{i_1 j_1} q_{i_1 j_2} q_{i_1 j_3} q_{j_1 j_2} q_{j_1 j_3} q_{j_2 j_3}$$

$$+ \sum_{j_1} \sum_{j_2} \sum_{j_3} q_{j_1 j_1} q^2_{j_1 j_2} q^2_{j_1 j_3} q_{j_2 j_3}$$

$$= - \left(\frac{1}{n}\right) \sum_{i_1 \neq j_1} \sum_{j_2} \sum_{j_3} q_{i_1 j_2} q_{i_1 j_3} q_{j_1 j_2} q_{j_1 j_3} q_{j_2 j_3} \tag{3.14}$$

$$+ \left(\frac{n-1}{n}\right) \sum_{j_1} \sum_{j_2} \sum_{j_3} q^2_{j_1 j_2} q^2_{j_1 j_3} q_{j_2 j_3}$$

$$= - \left(\frac{1}{n}\right) \sum_{i_1} \sum_{j_1} \sum_{j_2} \sum_{j_3} q_{i_1 j_2} q_{i_1 j_3} q_{j_1 j_2} q_{j_1 j_3} q_{j_2 j_3}$$

$$+ \sum_{j_1} \sum_{j_2} \sum_{j_3} q^2_{j_1 j_2} q^2_{j_1 j_3} q_{j_2 j_3}$$

$$= \left(\frac{1}{n}\right) \sum_{j_1} \sum_{j_2} \sum_{j_3} q^2_{j_2 j_3} q_{j_1 j_2} q_{j_1 j_3}$$

$$+ \sum_{j_1} \sum_{j_2} \sum_{j_3} q^2_{j_1 j_2} q^2_{j_1 j_3} q_{j_2 j_3}$$

$$= \left(\frac{1}{n}\right) \sum_{j_2} \sum_{j_3} q^3_{j_2 j_3} + \sum_{j_1} \sum_{j_2} \sum_{j_3} q^2_{j_1 j_2} q^2_{j_1 j_3} q_{j_2 j_3}$$

$$= \left(\frac{1}{n}\right)^3 (n-1)[(n-1)^2 - 1] + \sum_{j_1 \neq j_2} \sum_{j_3} q^2_{j_1 j_2} q^2_{j_1 j_3} q_{j_2 j_3}$$

$$+ \sum_{j_2} \sum_{j_3} q^2_{j_2 j_2} q^3_{j_2 j_3}$$

$$= - \left(\frac{1}{n}\right)^3 (n-1)[(n-1)^2 - 1] + \left(\frac{1}{n}\right)^2 \sum_{j_1 \neq j_2} \sum_{j_3} q^2_{j_1 j_3} q_{j_2 j_3}$$

$$+ \left(\frac{n-1}{n}\right)^2 \left(\frac{1}{n}\right)^2 (n-1)[(n-1)^2 - 1]$$

$$= \left[\left(\frac{n-1}{n}\right) - \frac{1}{n}\right] \left(\frac{1}{n}\right)^2 (n-1)[(n-1)^2 - 1] + \left(\frac{1}{n}\right)^2 \sum_{j_1} \sum_{j_2} \sum_{j_3} q_{j_1 j_3} q_{j_2 j_3}$$

$$- \left(\frac{1}{n}\right)^2 \sum_{j_2} \sum_{j_3} q_{j_2 j_3}$$

$$= \left[\left(\frac{n-1}{n}\right)^2 - \frac{1}{n} - \left(\frac{1}{n}\right)^2\right] \left(\frac{1}{n}\right)^2 (n-1)[(n-1)^2 - 1]$$

$$= \left(\frac{1}{n}\right)^3 (n-3)(n-1)[(n-1)^2 - 1] .$$

Other functions of the $\{q_{ij}\}$ are evaluated by similar manipulations.

The symmetric functions of the $\{q_{ij}\}$ for the $b \times t$ cross classification are immediately available from the corresponding functions for the simple random sample. The residual corresponding to an element in row i and column j can be written as

$$y_{ij} - \frac{1}{t} \sum_{j'} y_{ij'} - \frac{1}{b} \sum_{i'} y_{i'j} + \frac{1}{bt} \sum_{i'j'} y_{i'j'} \cdot$$

It follows that the covariance between the residual corresponding to row i, column j and row i' and column j' is equal to

$$\delta_{ii'} \delta_{jj'} - \frac{1}{t} \delta_{ii'} - \frac{1}{b} \delta_{jj'} + \frac{1}{bt} = (\delta_{ii'} - \frac{1}{b})(\delta_{jj'} - \frac{1}{t})$$

where $\delta_{hk} = 1$ if $h = k$ and zero otherwise. The matrix Q consisting of the $\{q_{ij}\}$ can then be written in the form

$$(I_{(b)} - \frac{1}{b} J_{(b)}) \otimes (I_{(t)} - \frac{1}{t} J_{(t)})$$

where $I_{(b)}$ and $I_{(t)}$ are the $b \times b$ and $t \times t$ identity matrices, $J_{(b)}$ and $J_{(t)}$ are the $b \times b$ and $t \times t$ matrices of ones and $\otimes$ denotes the direct product of the 2 matrices. If one thinks of Q as a scalar times the variance-covariance matrix of the residuals it is immediately clear that each element q_{ij} will consist of the product of a term depending on the row classification of the 2 residuals and a term depending on the column classification. Consequently the symmetric functions of the $\{q_{ij}\}$ will factor into the product of the symmetric functions which can be evaluated separately. Each will have the same form or the corresponding function for the simple random sample with n replaced by b and t respectively.

The extension to n-way arrays of values is immediately obvious and will not be discussed. Unfortunately this simplicity does not exist for designs which do not possess a balanced complete structure. The problem is that it is difficult to write an explicit function for the elements of the Q matrix. The difficulties are well illustrated by the Latin square design. The q_{ij} of the matrix Q is equal to a constant times the correlation between the residuals corresponding to observation i and observation j. This correlation depends not only on the relative row and column positions of the observations but also on whether they are subject to identical treatments. This latter, however, is a function of the particular Latin square selected when the experiment was performed.

For the balanced incomplete block design the elements of Q are functions of the relationships among the observed values. Consequently it was not possible to obtain manageable expressions for the elements of Q.

D. Numerical Example

For the simple random sample of size n (3.12) can be put into the form

$$
\begin{aligned}
E[g_1^*] = \frac{n}{(n-1)(n-2)\,\mu^{3/2}} \Bigg\{ & \frac{(n-1)(n-2)}{n} K_3 \qquad (3.15) \\
& - \left(\frac{3}{2}\right)\Big[(n-1)\ K_5 + 6n\ K_3K_2\Big]\left[\frac{(n-1)^2 - 1}{n^3\ K_2}\right] \\
& + \left(\frac{15}{8}\right)\Big[(n-1)^2\ K_7 + 16n(n-1)\ K_5K_2 + n(n-1)^2\ K_4K_3 + 12n(2n-3)\ K_4K_3 \\
& + 2n^2(n+23)\ K_3K_2^2\Big]\left[\frac{(n-2)}{n^3(n-1)\ K_2^2}\right] \\
& - \left(\frac{35}{16}\right)\Big[(n-1)^3\ K_9 + 30n(n-1)^2\ K_7K_2 \\
& + \Big(n(n+17)(n-1)^2 + 44n(n-1)\Big)\ K_6K_3 \\
& + \Big(3n(n+15)(n-1)^2 + 36n(n-1)(n-2) + 24(n-1)^2 + 24\Big)\ K_5K_4 \\
& + 6n^2(n-1)(n+29)\ K_5K_2^2 + 4n^2(n^2 + 21n - 52)\ K_3^3 \\
& + 24n^3(n+18)\ K_3K_2^3\Big]\left[\frac{(n-2)}{n^4(n-1)^2\ K_2^3}\right]\Bigg\} .
\end{aligned}
$$

From the form of (3.15) it is clear that as n becomes very large the measure of skewness will approach $K_3/K_2^{3/2}$. However, smaller n values will tend to reduce the measure of skewness. Since the last term of (3.15) is only of order n^{-2} rather large values of n will be needed in order to provide a reasonable approximation to $E[g_1]$ for distributions in which the higher cumulants become large. It is immediately obvious from the form of (3.15) that the expected value of g_1^* will be zero for a symmetrical distribution. As an example of the expected value of g_1^* when the errors are not symmetrically distributed, (3.15) will be evaluated for a simple random sample of 100 observations with errors drawn from a gamma distribution with parameters r and λ.

The characteristic function of this distribution is $(1 - \frac{it}{\lambda})^{-r}$ and the cumulant generating function is

$$- r \ln (1 - \frac{it}{\lambda}) = r \frac{it}{\lambda} + \frac{r}{2}\left(\frac{it}{\lambda}\right)^2 + \frac{r}{3}\left(\frac{it}{\lambda}\right)^3 + \frac{r}{4}\left(\frac{it}{\lambda}\right)^4 + \dots \quad . \tag{3.16}$$

The m'th cumulant can be obtained immediately as the coefficient of $\frac{(it)^m}{m!}$.

Consequently:

$$K_2 = \frac{r}{\lambda^2}$$

$$K_3 = \frac{2r}{\lambda^3}$$

$$K_4 = \frac{3!r}{\lambda^4}$$

$$K_5 = \frac{4!r}{\lambda^5}$$

$$K_6 = \frac{5!r}{\lambda^6} .$$

$$K_7 = \frac{6!r}{\lambda^7}$$

$$K_9 = \frac{8!r}{\lambda^9} .$$

For a sample of 100 one obtains

$$E[g_1^*] = 1.90r^{-1/2} - 1.50r^{-3/2} - .21r^{-5/2} - .09r^{-7/2} . \tag{3.17}$$

The negative terms in the expression are largely due to the effects of the large values of K_5, K_6, K_7 and K_9 . It is interesting to note that as r becomes large (3.17) will approach zero. This corresponds to the fact that the gamma distribution tends towards symmetry as r increases.

When sampling from the normal Fisher (1958) has shown that the variance of g_1 is

$$\frac{6n(n-1)}{(n-2)(n+1)(n+3)}$$

For a sample of 100 this is equal to .058 .

A measure of the power of a test based on g_1 can be obtained by evaluating (3.17) for several values of r and comparing with the standard deviation of the statistic under the ideal conditions of NID(0, σ^2) errors. For r = 4, $E[g_1^*]$ is .65 or 2 1/2 times the standard deviation for samples from the normal. For r = 9, the expected value becomes .58 and for r = 25 it becomes .37 , indicating that for samples of one hundre, one probably has a reasonable chance of rejecting populations whose shape is similar to the gamma with r = 9 . Larger samples will be needed if sensitivity to gamma distributions with larger r values is desired.

E. Higher Moments of the Measure of Skewness

The methods described in the previous sections can be extended to obtain higher moments. However, when this is done, certain difficulties arise. The most important of these is the excessive number of terms which appear in various expressions and must be considered. An attempt to examine the variance of g_1 for non-normal errors was abandoned because of this reason.

The approach used was to obtain an approximate expression for $E[g_1^2]$ and use it in conjunction with the expression for the mean of g_1 to obtain the variance. Following (3.2), $[g_1]^2$ can be written as

$$\frac{(\sum_{ij} q_{ij}^2)^3 \; (\sum_i z_i^3)^2}{(\sum_{ij} q_{ij}^3)^2 \; (\nu\mu_2)^3} \left[1 + \frac{\delta}{\nu\mu_2}\right]^{-3} , \tag{3.18}$$

and this expression approximated by

$$\frac{(\Sigma\, q_{ij}^2)^3 \; (\Sigma\, z_i^3)^2}{(\Sigma\, q_{ij}^3)^2 \; (\nu\mu_2)^3} \left[1 - 3\,\frac{\delta}{\nu\mu_2} + 6\left(\frac{\delta}{\nu\mu_2}\right)^2 - 10\left(\frac{\delta}{\nu\mu_2}\right)^3\right] . \tag{3.19}$$

Writing δ as $\sum z_i^2 - \nu\mu_2$ and collecting terms leads to

$$\frac{(\Sigma\, q_{ij}^2)^3 \; (\Sigma\, z_i^3)^2}{(\Sigma\, q_{ij}^3)^2 \; (\nu\mu_2)^3} \left[20 - 45\left(\frac{\Sigma\, z_i^2}{\nu\mu_2}\right) + 36\left(\frac{\Sigma\, z_i^2}{\nu\mu_2}\right)^2 - 10\left(\frac{\Sigma\, z_i^2}{\nu\mu_2}\right)^3\right] \tag{3.20}$$

In order to compute the expected value of expression (3.20), the expected value of the 4 terms, $(\sum z_i^3)^2$, $(\sum z_i^3)^2 (\sum z_j^2)$, $(\sum z_i^3)^2 (\sum z_j^2)^2$ and $(\sum z_i^3)^2 (\sum z_j^2)^3$ must be evaluated. The first two of these quantities can be obtained easily.

$$E[\sum z_i^3 , \sum z_j^3] = (111222)\ K_6$$

$$+ [6(1112)(22) + 9(1122)(12)]\ K_4K_2$$

$$+ [(111)(222) + 9(112)(122)\ K_3^2$$

$$+ [9(11)(12)(22) + 6(12)(12)(12)]\ K_2^3 .$$

$$E[\sum z_i^3 , \sum z_j^3 , \sum z_h^2] = (11122233)\ K_8$$

$$+ [(111222)(33) + 12(111223)(23) + 6(111233)(22) + 9(112233)(12)]\ K_6K_2$$

$$+ [6(1112)(2233) + 2(1113)(2223) + 9(1122)(1233)$$

$$+ 18(1123)(1223)]\ K_4^2 .$$

$$+ [18(2233)(11)(12) + 12(2223)(11)(13) + 6(1222)(11)(33)$$

$$+ 18(1233)(12)(12) + 72(1223)(12)(13) + 9(1122)(12)(33)$$

$$+ 12(1222)(13)(13) + 9(1233)(11)(22) + 36(1223)(11)(23)$$

$$+ 18(1122)(13)(23)]\ K_4K_2^2$$

$$+ [(111)(222)(33) + 12(111)(223)(23) + 6(111)(233)(22) + 9(112)(122)(33)$$

$$+ 72(112)(123)(23) + 18(112)(133)(22) + 36(112)(223)(13)$$

$$+ 36(112)(233)(12) + 36(113)(123)(22) + 18(113)(223)(12)$$

$$+ 36(123)(123)(12)]\ K_3^2K_2$$

$$+ [2(111)(22233) + 18(112)(12233) + 12(113)(12223) + 18(123)(11223)$$

$$+ 6(133)(11222)]\ K_5K_3$$

$$+ [9(11)(12)(22)(33) + 36(11)(12)(23)(23) + 18(11)(13)(22)(23)$$

$$+ 6(12)(12)(12)(33) + 36(12)(12)(13)(23)]\ K_2^4 \ .$$

These expressions are written in full generality and can be reduced considerably by using the fact that $\sum_i q_{ij} = 0$ and $\sum_i q_{ij}^2 = \nu$. However, all terms and their coefficients must be evaluated in order to check that no term has been omitted.

The expected value of $[\sum z_i^3 , \sum z_j^3 , \sum z_h^2 , \sum z_n^2]$ was obtained but will not be reproduced here. It is sufficient to point out that the coefficient of K_8K_2 involved 5 terms, of K_7K_3 involved 7 terms, of K_6K_4 involved 10 terms, of $K_6K_2^2$ involved 18 terms, of $K_3^2K_2^2$ involved 58 terms, and so on.

The expected value of $(\sum z_i^3)^2\ (\sum z_j^2)^3$ was not obtained because of the excessively large number of terms involved. The magnitude of the number of terms can be surmised by considering the coefficient of $K_3^2K_2^3$ as an example. The problem is to enumerate the ways in which the collection of integers (111222334455) can be assigned to two sets of three and three sets of two. Enumeration shows that there are 32 distinct triples and 15 pairs. Sets of 2 triples and 3 pairs can be assembled in $\frac{32 \times 33}{2} \times \frac{15 \times 16 \times 17}{6} = 359{,}040$ distinct ways. These must then be examined in order to select those assignments which are valid, i.e., involve exactly 3 ones, 3 twos, 2 threes, 2 fours and 2 fives as well as to pick out those which are actually identical and only represent permutations of the digits.

IV. A MEASURE OF KURTOSIS WHEN ADDITIVITY HOLDS

A. Introduction

In Chapter II techniques for evaluating the moments of the residuals in terms of the moments of the distribution of the errors in the original observations were developed. In Chapter III these techniques were used to examine the behavior of the first scale-invarient shape coefficient of the error distribution, measuring skewness. The purpose of this chapter is to use the same techniques to examine a commonly used measure of kurtosis. The chapter will conclude with an example dealing with the case where the errors are independent samples from a mixture of two normal populations with unequal variances.

B. The Measure of Kurtosis

A commonly used measure of kurtosis is the statistic

$$g_2 = \left[\frac{\sum_i^n z_i^4}{s^4} - \frac{3\nu \sum_i^n q_{ii}^2}{\nu+2} \right] D^{-1} , \tag{4.1}$$

where

$$D = \sum_{ij}^n q_{ij}^4 - \frac{3\left(\sum_i^n q_{ii}^2\right)^2}{\nu(\nu+2)} , \tag{4.2}$$

$$\nu = \sum_{ij}^n q_{ij}^2$$

and

$$s^2 = \sum_{i}^{n} z_i^2/\nu \tag{4.3}$$

provided that $D \neq 0$.

$\{z_i\}$ and $\{q_{ij}\}$ are defined as in the previous chapters.

Since both D and the second term in (4.1) are functions of the design and are known without error, they can be ignored while developing an approximation for g_2 .

For the remainder of this chapter μ_m will be taken to mean the m'th moment about the mean for the errors in the observed values. Consequently (4.3) becomes the estimate of μ_2 and one can write

$$\delta = \sum_{i} z_i^2 - \nu\mu_2 \ . \tag{4.4}$$

It follows that

$$\frac{\left(\sum_{ij} q_{ij}^2\right)^2 \sum_{i} z_i^4}{\left(\sum_{i} z_i^2\right)^2} = \frac{\sum_{i} z_i^4}{(\mu_2)^2} \ [1 + \frac{\delta}{\nu\mu_2}]^{-2} \ . \tag{4.5}$$

The validity of this expression follows from the fact that μ_2 is a variance and hence non-zero and ν can be assumed to be positive integer. The right hand term in (4.5) can now be expanded to yield

$$\frac{\Sigma z_i^4}{(\mu_2)^2} \ [1 + \frac{\delta}{\nu\mu_2}]^{-2}$$

$$= \frac{\Sigma z_i^4}{(\mu_2)^2} \left[1 - \frac{2\delta}{\nu\mu_2} + \frac{3\delta^2}{(\nu\mu_2)^3} - \frac{4\delta^3}{(\nu\mu_2)^3} + \dots\right] . \tag{4.6}$$

This series expansion can be extended to any number of terms. The expected value of $\sum z_i^4 \delta^m$ will be of order n^2 while ν is of the same order of magnitude as n . Consequently it is always possible to find a value of n sufficiently large that the last terms retained in the series will contribute very little. For the purpose of the examination in this chapter the expansion will be terminated with the cubic term in δ .

One can write

$$E[g_2^*] = \frac{1}{(\mu_2)^2 D} \left\{ E[\sum z_i^4] - \frac{2E[\Sigma z_i^4 (\Sigma z_j^2 - \nu\mu_2)]}{\nu\mu_2} + \frac{3E[\Sigma z_i^4 (\Sigma z_j^2 - \nu\mu_2)^2]}{(\nu\mu_2)^2} - \frac{4E[\Sigma z_i^4 (\Sigma z_j^2 - \nu\mu_2)^3]}{(\nu\mu_2)^3} - C \right\}, \tag{4.7}$$

where

$$C = \frac{3\nu(\Sigma q_{ii}^2)}{\nu+2} .$$

For the case where the residuals are assumed to have zero sum and variances equal to ν/n one can write

$$C = \frac{3\nu^3}{n(\nu+2)} \tag{4.8}$$

and

$$D = \sum q_{ij}^4 - \frac{3\nu^3}{n^2(\nu+2)} . \tag{4.9}$$

(4.7) can also be written in the form

$$g_2^* = \frac{1}{(\mu_2)^2 D} \left\{ 10E[\sum z_i^4] - \frac{20E[(\Sigma z_i^4)(\Sigma z_j^2)]}{\nu\mu_2} \right. \tag{4.10}$$

$$\left. + \frac{15E[(\Sigma z_i^4)(\Sigma z_j^2)^2]}{(\nu\mu_2)^2} - \frac{3E[(\Sigma z_i^4)(\Sigma z_j^2)^3]}{(\nu\mu_2)^3} - C \right\} .$$

It is clear from (4.7) and (4.10) that the expected values of four different polynomial functions of the residuals must be evaluated. This can be done using the techniques developed in Chapter II. However, calculations will be performed only for the special case discussed in the next section.

C. Sampling from Symmetric Distributions

The expected values of the residuals required in (4.7) and (4.10) will now be evaluated for the case where the e_i are independent random samples from a symmetric population. Attention will eventually be restricted to a population consisting of a fraction p from $N(0, \sigma_1^2)$ and (1-p) from $N(0, \sigma_2^2)$. The rational for choosing this distribution is that it is often appropriate to assume that a fraction of the observations, though unbiased, are subject to errors of considerably larger magnitude than the remainder.

Using the system of notation developed in Chapter III we have:

$$E[\sum z_i^4] = (1111)\ K_4 + 3(11)(11)\ K_2^2 . \tag{4.11}$$

$$E[(\sum z_i^4)(\sum z_j^2)] = (111122)\ K_6 \tag{4.12}$$

$+ [(1111)(22) + 8(1112)(12) + 6(1122)(11)]\ K_4K_2$

$+ [3(11)(11)(22) + 12(11)(12)(12)]\ K_2^3$.

$$E[(\sum z_i^4)(\sum z_j^2)(\sum z_h^2)] = (11112233)\ K_8 \tag{4.13}$$

$+ [2(111122)(33) + 4(111123)(23) + 6(112233)(11) + 16(111223)(13)]\ K_6K_2$

$+ [(1111)(2233) + 16(1112)(1233) + 6(1122)(1133) + 12(1123)(1123)]\ K_4^2$

$+ [(1111)(22)(33) + 2(1111)(23)(23) + 16(1112)(12)(33) + 32(1112)(13)(23)$

$+ 12(1122)(11)(33) + 24(1122)(13)(13) + 48(1223)(11)(13) + 3(2233)(11)(11)$

$+ 24(1123)(11)(23) + 48(1123)(12)(13)]\ K_4K_2^2$

$+ [3(11)(11)(22)(33) + 24(11)(12)(12)(33) + 6(11)(11)(23)(23)$

$+ 48(11)(12)(13)(23) + 24(12)(12)(13)(13)]\ K_2^4$.

$$E[(\sum z_i^4)(\sum z_j^2)(\sum z_h^2)(\sum z_k^2)] \tag{4.14}$$

$= (1111223344)\ K_{10} + [3(11112233)(44) + 12(11112234)(34)$

$+ 24(11123344)(12) + 6(11223344)(11)]\ K_8K_2$

$+ [3(111122)(3344) + 12(111123)(2344) + 48(111223)(1344)$

$+ 32(111234)(1234) + 18(112233)(1144) + 72(112344)(1123)$

$+ 24(122334)(1114) + (223344)(1111)]\ K_6K_4$

$+ [3(111122)(33)(44) + 6(111122)(34)(34) + 12(111123)(23)(44)$

$+ 24(111123)(24)(34) + 48(111223)(13)(44) + 96(111223)(14)(34)$

$+ 96(111234)(12)(34) + 18(112233)(11)(44) + 36(112233)(14)(14)$

$+ 72(122334)(11)(14) + 3(223344)(11)(11) + 72(112234)(11)(34)$

$$+ 144(112234)(13)(14)]\ K_6K_2^2$$

$$+ [3(1111)(2233)(44) + 12(1111)(2234)(34) + 48(1112)(1233)(44)$$

$$+ 96(1112)(1334)(24) + 96(1112)(1234)(34) + 24(1112)(3344)(12)$$

$$+ 96(1112)(2334)(14) + 144(1122)(1334)(14) + 18(1122)(3344)(11)$$

$$+ 18(1122)(1133)(44) + 288(1123)(1244)(13) + 288(1123)(1234)(14)$$

$$+ 72(1123)(2344)(11) + 144(1123)(1124)(34) + 72(1223)(1344)(11)$$

$$+ 36(1123)(1123)(44) + 48(1234)(1234)(11) + 72(1122)(1134)(34)]\ K_4^2K_2$$

$$+ [(1111)(22)(33)(44) + 6(1111)(22)(34)(34) + 8(1111)(23)(24)(34)$$

$$+ 48(1112)(12)(34)(34) + 24(1112)(12)(33)(44) + 96(1112)(13)(23)(44)$$

$$+ 192(1112)(13)(24)(34) + 18(1122)(11)(33)(44) + 36(1122)(11)(34)(34)$$

$$+ 72(1122)(13)(13)(44) + 144(1122)(13)(14)(34) + 124(1223)(11)(13)(44)$$

$$+ 288(1223)(11)(14)(34) + 288(1223)(13)(14)(14) + 9(2233)(11)(11)(44)$$

$$+ 36(2233)(11)(14)(14) + 36(2234)(11)(11)(34) + 144(2234)(11)(13)(14)$$

$$+ 288(1234)(11)(12)(34) + 192(1234)(12)(13)(14) + 72(1123)(11)(23)(44)$$

$$+ 144(1123)(11)(24)(34) + 144(1123)(12)(13)(44) + 576(1123)(12)(14)(34)$$

$$+ 144(1123)(14)(14)(23)]\ K_4K_2^3$$

$$+ [3(11)(11)(22)(33)(44) + 18(11)(11)(22)(34)(34) + 24(11)(11)(23)(24)(34)$$

$$+ 36(11)(12)(12)(33)(44) + 72(11)(12)(12)(34)(34) + 144(11)(12)(13)(23)(44)$$

$$+ 288(11)(12)(13)(24)(34) + 72(12)(12)(13)(13)(44)$$

$$+ 288(12)(12)(13)(14)(34)]\ K_2^5\ .$$

If all the residuals are assumed to have zero sum and variance ν/n then the foregoing can be simplified considerably. (4.11) can be written as

$$E[\sum z_i^4] = (1111)\ K_4 + \frac{3\nu^2}{n} K_2^2\ . \tag{4.15}$$

(4.12) can be written as

$$E[(\sum z_i^4)(\sum z_j^2)] = \frac{\nu}{n}\ (1111)\ K_6 \tag{4.16}$$

$$+ [8(1111) + \nu(1111) + 6\left(\frac{\nu}{n}\right)^2 \nu]\ K_4K_2$$

$$+ [3\left(\frac{\nu}{n}\right) \nu^2 + 12\left(\frac{\nu}{n}\right) \nu]\ K_2^3\ .$$

(4.13) can be written as

$$E[(\sum z_i^4)(\sum z_j^2)(\sum z_h^2)] \tag{4.17}$$

$$= \left(\frac{\nu}{n}\right)^2 (1111)\ K_8 + [2\left(\frac{\nu}{n}\right) \nu(1111) + 20\left(\frac{\nu}{n}\right) (1111) + 6\left(\frac{\nu}{n}\right)^3 \nu]\ K_6K_2$$

$$+ [\left(\frac{\nu}{n}\right) \nu(1111) + 16\left(\frac{\nu}{n}\right) (1111) + 6\left(\frac{\nu}{n}\right)^3 \nu + 12(1123)(1123)]\ K_4^2$$

$$+ [\nu^2 (1111) + 18\nu(1111) + 80(1111) + 15\left(\frac{\nu}{n}\right)^2 \nu^2 + 96\left(\frac{\nu}{n}\right)^2 \nu]\ K_4K_2^2$$

$$+ [3\left(\frac{\nu}{n}\right) \nu^3 + 30\left(\frac{\nu}{n}\right) \nu^2 + 72\left(\frac{\nu}{n}\right) \nu]\ K_2^4\ .$$

(4.14) can be written as

$$E[(\sum z_i^4)(\sum z_j^2)(\sum z_h^2)(\sum z_k^2)] \tag{4.18}$$

$$= \left(\frac{\nu}{n}\right)^3 (1111)\ K_{10} + [3\left(\frac{\nu}{n}\right)^2 \nu(1111) + 36\left(\frac{\nu}{n}\right)^2 (1111) + 6\left(\frac{\nu}{n}\right)^4 \nu]\ K_8K_2$$

$$+ [4\left(\frac{\nu}{n}\right)^2 \nu(1111) + 72\left(\frac{\nu}{n}\right)^2 (1111) + 12\left(\frac{\nu}{n}\right) \nu(1111)$$

$$+ 72\left(\frac{\nu}{n}\right) (1123)(1123) + 32(111234)(1234) + 18\left(\frac{\nu}{n}\right)^4 \nu]\ K_6K_4$$

$$+ [3\left(\frac{\nu}{n}\right) \nu^2(1111) + 66\left(\frac{\nu}{n}\right) \nu(1111) + 360\left(\frac{\nu}{n}\right) (1111)$$

$$+ 180\left(\frac{\nu}{n}\right)^3 \nu + 21\left(\frac{\nu}{n}\right)^3 \nu^2]\ K_6K_2^2$$

$$+ [84\left(\frac{\nu}{n}\right) \nu(1111) + 624\left(\frac{\nu}{n}\right) (1111) + 3\left(\frac{\nu}{n}\right) \nu^2(1111)$$

$$+ 432(1123)(1123) + 36\nu(1123)(1123)$$

$$+ 360\left(\frac{\nu}{n}\right)^3 \nu + 36\left(\frac{\nu}{n}\right)^3 \nu^2]\ K_4^2K_2$$

$$+ [\nu^3(1111) + 30\nu^2(1111) + 296\nu(1111) + 960(1111)$$

$$+ 27\left(\frac{\nu}{n}\right)^2 \nu^3 + 396\left(\frac{\nu}{n}\right)^2 \nu^2 + 1440\left(\frac{\nu}{n}\right) \nu]\ K_4K_2^3$$

$$+ [3\left(\frac{\nu}{n}\right) \nu^4 + 54\left(\frac{\nu}{n}\right) \nu^3 + 312\left(\frac{\nu}{n}\right) \nu^2 + 576\left(\frac{\nu}{n}\right) \nu]\ K_2^5 .$$

In order to apply these formulas to a specific design the four quantities ν, (1111), (1123)(1123) and (111234)(1234) must be evaluated. For the simple random sample of size n ,

$$\nu = n - 1 ,$$

$$(1111) = \frac{(n-1)[(n-1)^3 + 1]}{n^3} , \tag{4.19}$$

$$(1123)(1123) = \frac{(n-1)^4(n-2) + (n-1)(n^2-2)}{n^4} , \tag{4.20}$$

and

$$(111234)(1234) = \frac{(n-1)^6}{n^5} + \frac{(n-1)}{n^5} - \frac{(n-1)^2(n-2)^2}{n^4} \tag{4.21}$$

For a two-way cross classification consisting of r rows and c columns,

$$\nu = (r-1)(c-1) ,$$

$$(1111) = \left(\frac{(r-1)[(r-1)^3 + 1]}{r^3}\right)\left(\frac{(c-1)[(c-1)^3 + 1]}{c^3}\right) , \tag{4.22}$$

$$(1123)(1123) = \left(\frac{(r-1)^4(r-2) + (r-1)(r^2-2)}{r^4}\right) \times$$

$$\left(\frac{(c-1)^4(c-2) + (c-1)(c^2-2)}{c^4}\right) , \tag{4.23}$$

and

$$(111234)(1234) = \left(\frac{(r-1)^6 + (r-1) - r(r-1)^2(r-2)^2}{r^5}\right) \times$$

$$\left(\frac{(c-1)^6 + (c-1) - c(c-1)^2(c-2)^2}{c^5}\right) . \tag{4.24}$$

When samples are drawn from a mixture of two normal distributions, the first 10 cumulants can be obtained readily by evaluating the moments of the composite distribution and substituting into the formulas given by Kendall (1952). If a variate e is drawn from a $N(0, \sigma_1^2)$ population with probability p and from $N(0, \sigma_2^2)$ population 1-p then the first 10 moments are:

$$E[e] = 0$$

$$E[e^2] = p\sigma_1^2 + (1-p)\sigma_2^2 \cdot$$

$$E[e^3] = 0 .$$

$$E[e^4] = 3[p\sigma_1^4 + (1-p)\sigma_2^4] .$$

$$E[e^5] = 0 .$$

$$E[e^6] = 15[p\sigma_1^6 + (1-p)\sigma_2^6] .$$

$$E[e^7] = 0 .$$

$$E[e^8] = 105[p\sigma_1^8 = (1-p)\sigma_2^8] .$$

$$E[e^9] = 0$$

$$E[e^{10}] = 945[p\sigma_1^{10} + (1-p)\sigma_2^{10}] .$$

The first 5 non-zero cumulants are:

$$K_2 = p\sigma_1^2 + (1-p)\sigma_2^2 . \tag{4.25}$$

$$K_4 = 3p(1-p)(\sigma_1^2-\sigma_2^2)^2 . \tag{4.26}$$

$$K_6 = 15p(1-p)(1-2p)(\sigma_1^2-\sigma_2^2)^3 . \tag{4.27}$$

$$K_8 = 105p(1-p)(1-6p+6p^2)(\sigma_1^2-\sigma_2^2)^4 . \tag{4.28}$$

$$K_{10} = 945p(1-p)(1-2p)(1-12p+12p^2)(\sigma_1^2-\sigma_2^2)^5 \ . \tag{4.29}$$

D. Numerical Examples

Because of the complexity of the equations developed, it was decided to use numerical techniques to study the behavior of the approximation for g_2 for varying values of p, σ_2^2/σ_1^2 and n . For simplicity these computations will be performed for the simple random sample, though the extension to the two-way table is straight-forward.

Consider first the behavior of g_2^* for $\sigma_2/\sigma_1 = 2$ and n and p taking on the values 100, 200 and 300 and .95, .90 and .80 respectively. The results of the computations are given in Table 3.

Table 3. Values Obtained for the Approximation For the Measure of Kurtosis When $\sigma_2/\sigma_1 = 2$.

n	p = .95	p = .90	p = .80
100	.544	.975	1.175
200	.796	1.214	1.491
300	.859	1.298	1.567

The magnitude of these values can be judged by comparison with the mean and variance under ideal statistical conditions. The appropriate formulas are given by Anscombe (1961). The mean is always zero. For a simple random sample of size 100, the variance of g_2 is .224. For simple samples of size 200 and 300 the variances are .116 and .0782.

The influence of changes in the ratio σ_2/σ_1 can be observed in

Table 4. This table shows the values obtained from a simple random sample of 300, with p taking on the values .95, .90 and .80, while the ratio of standard derivations took on the values 1., 2., 3. and 4. The case for $\sigma_1 = \sigma_2$ was included in order to indicate the degree to which the approximate value agreed with the known exact value, zero.

Table 4. Values Obtained for the Approximation for The Measure of Kurtosis for Random Samples of 300.

σ_2/σ_1	1.	2.	3.	4.
p				
.95	-.002	.859	3.560	5.676
.90	-.002	1.298	4.424	7.118
.80	-.002	1.567	4.126	6.006

Under the ideal statistical conditions the standard deviation of the measure of kurtosis for the simple random sample of 300 is .28. The decrease evident in the lower right corner of the table may be due in part to a deficiency in the expansion. For example, for $n = 300$, $\sigma_1 = 1$, $\sigma_2 = 4$ and $p = .8$,

$$E[\sum z_i^4] = 46,274 ,$$

$$\frac{E[\Sigma\, z_i^4(\Sigma\, z_j^2 - \nu\mu_2)]}{\nu\mu_2} = 2725 ,$$

$$\frac{E[\Sigma\ z_i^4\ (\Sigma\ z_j^2 - \nu\mu_2)^2]}{(\nu\mu_2)^2} = 1581$$

and

$$\frac{E[\Sigma\ z_i^4\ (\Sigma\ z_j^2 - \nu\mu_2)^3]}{(\nu\mu_2)^3} = 669\ .$$

This sequence suggests that if the next term in the series expansion were computed, it could influence the results to a minor extent.

V. SOME RESULTS UNDER AN ALTERNATIVE MODEL

A. Introduction

In the previous chapter the properties of the residuals were examined under an assumed linear model incorporating an error term from some infinite population. In this chapter the emphasis will shift to an examination of the residuals for a randomized block design when additivity holds. Kempthorne (1952) has developed the appropriate model for the case where t treatments are applied at random to the t plots in each of b blocks, with the restriction that there is exactly one treatment per plot, each treatment occurs in every block, and additivity holds. The model is

$$y_{ij} = \mu + \beta_i + \tau_j + \sum_{k=1}^{t} \delta_{ij}^{k} \varepsilon_{ik} , \tag{5.1}$$

where β_i is the effect due to block i , τ_j is the effect due to block j , ε_{ik} is the effect due to plot k in block i and δ_{ij}^{k} is a random variable equal to unity if treatment k occurs on plot j in block i and zero otherwise. Also $\sum_i \beta_i = 0$, $\sum_j \tau_j = 0$ and $\sum_k \varepsilon_{ik} = 0$ for all i .

The model can be extended easily to the case where the treatments are applied to a random sample of t plots from a population of n in each block. Increasing n under this formulation causes a decrease in the correlation between errors in the same block, with the limiting case being independence. The model now becomes

$$y_{ij} = \mu + \beta_i + \tau_j + \sum_{k=1}^{n} \delta_{ij}^{k} \varepsilon_{ik} \tag{5.2}$$

The only random elements in the model are the $\{\delta_{ij}^{k}\}$. Their joint distribution is completely specified by the randomization procedure. The first and second moments are

$$E[\delta_{ij}^{k}] = \frac{1}{n} \quad , \tag{5.3}$$

$$E[\delta_{ij}^{k}\ \delta_{i'j'}^{k'}] \begin{cases} = \frac{1}{n} \text{ if } i=i' \ , \ j=j' + k=k' \\ = E[\delta_{ij}^{k}]\ E[\delta_{i'j'}^{k}] \quad \text{if } i\neq i' \\ = 0 \quad \text{if } i=i' \ , \ j=j' \ , \ k\neq k' \\ = 0 \quad \text{if } i=i' \ , \ j\neq j' \ , \ k=k' \\ = \frac{1}{n(n-1)} \quad \text{if } i=i' \ , \ j\neq j' \ , \ k\neq k' \ . \end{cases} \tag{5.4}$$

It follows that

$$E[y_{ij}] = E[\mu + \beta_i + \tau_j + \sum_{k=1}^{n} \delta_{ij}^{k}\ \varepsilon_{ik}] = \mu + \beta_i + \tau_j \ . \tag{5.5}$$

Suitable estimates of μ , β_i and τ_j are $y_{..}$, $y_{i.} - y_{..}$ and $y_{.j} - y_{..}$ respectively for all values of i and j. These can be combined to provide an improved estimate of the response due to treatment j on block i. This estimate is

$$\hat{y}_{ij} = y_{i.} + y_{.j} - y_{..} \ . \tag{5.6}$$

This estimate is unbiased since

$$E[\hat{y}_{ij}] = \frac{1}{b} \sum_{i=1}^{b} E[y_{i\cdot}] + \frac{1}{t} \sum_{j=1}^{t} E[y_{\cdot j}] - \frac{1}{bt} \sum_{i=1}^{b} \sum_{j=1}^{t} E[y_{ij}] \tag{5.7}$$

$$= \mu + \beta_i + \tau_j \ .$$

The residuals z_{ij} can now be defined by the equations

$$z_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - y_{i\cdot} - y_{\cdot j} + y_{\cdot\cdot} \tag{5.8}$$

for i=1, 2, ..., b and j=1, 2, ..., t . It follows immediately that these residuals have zero expectation.

B. Variances and Covariances

Equation (5.8) can be rewritten as

$$z_{ij} = \frac{(b-1)(t-1)}{bt} y_{ij} - \sum_{k\neq j} \frac{b-1}{bt} y_{ik} - \sum_{h\neq i} \frac{t-1}{bt} y_{hj}$$

$$+ \frac{1}{bt} \sum_{h\neq i} \sum_{k\neq j} y_{hk} \ . \tag{5.9}$$

$$= \sum_k \frac{(b-1)}{bt} (t\ \delta_{jk} - 1)\ y_{ik} - \sum_{h\neq 1} \sum_k \frac{1}{bt} (t\ \delta_{jk} - 1)\ y_{hk}$$

$$= \sum_k a_{jk}\ y_{ik} - \frac{1}{b-1} \sum_{h\neq i} \sum_k a_{jk}\ y_{hk}$$

where

$$a_{ij} = \frac{(b-1)}{bt} (t\ \delta_{ij} - 1)$$

and

$$\delta_{ij} \begin{cases} = 1 & \text{if } i=j \\ = 0 & \text{if } i\neq j \ . \end{cases}$$

Substituting (5.2) into (5.9) leads to the equation

$$z_{ij} = \sum_{\ell=1}^{t} a_{j\ell} \left[\sum_{k=1}^{n} \delta_{i\ell}^{k} \varepsilon_{ik} - \frac{1}{b-1} \sum_{h\neq i}^{b} \sum_{k=1}^{n} \delta_{h\ell}^{k} \varepsilon_{hk} \right] . \tag{5.10}$$

Since the $\{a_{ij}\}$ are known numbers determined by the size of the experiment, it is obvious that $E[z_{ij}] = 0$.

Consequently all moments can be derived by obtaining the expected values of the appropriate products of residuals.

To simplify the presentation let

$$e_{i\ell} = \sum_{k=1}^{n} \delta_{i\ell}^{k} \varepsilon_{ik}$$

for all i and ℓ . It follows that

$$z_{ij} = \sum_{\ell}^{t} a_{j\ell} \left[e_{i\ell} - \frac{1}{b-1} \sum_{h\neq 1}^{b} e_{h\ell} \right] . \tag{5.11}$$

The properties of the $\{\delta_{ij}^{k}\}$ stated in (5.4) immediately imply that

$$E[e_{ij}\, e_{i'j'}] = E\left[\left(\sum_{k=1}^{n} \delta_{ij}^{k} \varepsilon_{ik}\right)\left(\sum_{k'=1}^{n} \delta_{i'j'}^{k'} \varepsilon_{i'k'}\right)\right] \tag{5.12}$$

$$= E\left[\sum_{k} \delta_{ij}^{k} \delta_{i'j'}^{k} \varepsilon_{ik} \varepsilon_{i'k} + \sum_{k\neq k'} \delta_{ij}^{k} \delta_{i'j'}^{k} \varepsilon_{ik} \varepsilon_{i'k'}\right]$$

$$= \delta_{ii'} \delta_{jj'} \left(\frac{1}{n}\right) \sum_{k} \varepsilon_{ik}^{2} - \delta_{ii'} (1-\delta_{jj'}) \left(\frac{1}{n(n-1)}\right) \sum_{k} \varepsilon_{ik}^{2}$$

$$= \delta_{ii'}\ \delta_{jj'}\ (\tfrac{n-1}{n})\ \sigma_i^2 - \delta_{ii'}\ (1-\delta_{jj'})(\tfrac{1}{n})\ \sigma_i^2$$

where $\sigma_i^2 = \frac{1}{n-1} \sum_k^n \varepsilon_{ik}^2$. For the remainder of this section it will be assumed that the within block variances are equal and

$$\frac{1}{n-1} \sum_k^n \varepsilon_{ik}^2 = \sigma^2 \text{ for all } i \text{ .}$$

This result is an immediate contrast to the infinite error model discussed in the previous chapters in which the errors associated with two observations were uncorrelated. However, the covariance structure of the residuals is identical under the two models. This can be shown as follows:

For observations in the same block

$$E[z_{ij}\ z_{ij'}] = E\Big[(\sum_{\ell} a_{j\ell}\ [e_{i\ell} - (\tfrac{1}{b-1}) \sum_{h\neq i} e_{h\ell}])(\sum_{\ell'} a_{j'\ell'} - (\tfrac{1}{b-1}) \sum_{h\neq i} e_{h'\ell'}])\Big]$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell}\ (E[e_{i\ell}^2 + (\tfrac{1}{b-1})^2 \sum_{h\neq i} e_{h\ell}^2)$$

$$+ \sum_{\substack{\ell,\ell' \\ \ell\neq\ell'}} a_{j\ell}\ a_{j'\ell'} \Big(E[e_{i\ell}\ e_{i\ell'} + (\tfrac{1}{b-1})^2 \sum_{h\neq i} e_{h\ell}\ e_{h\ell'}]\Big)$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell} \Big((\tfrac{n-1}{n})\ \sigma^2 + (\tfrac{1}{b-1})\ \tfrac{n-1}{n}\ \sigma^2\Big)$$

$$+ \sum_{\substack{\ell,\ell' \\ \ell\neq\ell'}} a_{j\ell}\ a_{j'\ell'} \Big(- (\tfrac{1}{n})\ \sigma^2 - (\tfrac{1}{n})(\tfrac{1}{b-1})\ \sigma^2\Big)$$

$$= a_{jj'}\ \sigma^2 \text{ .}$$

For observations in different blocks

$$E[z_{ij}\ z_{i'j'}] = E\Big[\big(\sum_{\ell} a_{j\ell}\ [e_{i\ell} - \left(\tfrac{1}{b-1}\right) \sum_{h\neq i} e_{h\ell}]\big)$$

$$\big(\sum_{\ell'} a_{j'\ell'}\ [e_{i'\ell'} - \left(\tfrac{1}{b-1}\right) \sum_{h'\neq i'} e_{h'\ell'}]\big)\Big]$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell}(E[\ -\left(\tfrac{1}{b-1}\right) e_{i\ell}^2 - \left(\tfrac{1}{b-1}\right) e_{i'\ell}^2 + \left(\tfrac{1}{b-1}\right)^2 \sum_{\substack{h\neq i\\ \neq i'}} e_{h\ell}^2])$$

$$+ \sum_{\substack{\ell,\ell'\\ \ell\neq\ell'}} a_{j\ell}\ a_{j'\ell'} \Big(E[\ -\left(\tfrac{1}{b-1}\right) e_{i\ell}\ e_{i\ell'} - \left(\tfrac{1}{b-1}\right) e_{i'\ell}\ e_{i'\ell'}$$

$$+ \left(\tfrac{1}{b-1}\right)^2 \sum_{\substack{h\neq i\\ \neq i'}} e_{h\ell}\ e_{h\ell'}]\Big)$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell}\ (\ -\left(\tfrac{2}{b-1}\right)\left(\tfrac{n-1}{n}\right) \sigma^2 + \left(\tfrac{1}{b-1}\right)^2 (b-2)\left(\tfrac{n-1}{n}\right) \sigma^2)$$

$$+ \sum_{\substack{\ell,\ell'\\ \ell\neq\ell'}} a_{j\ell}\ a_{j'\ell'}\ (\ \left(\tfrac{2}{b-1}\right)\left(\tfrac{1}{n}\right) \sigma^2 - \left(\tfrac{1}{b-1}\right)^2 (b-2)\left(\tfrac{1}{n}\right) \sigma^2)$$

$$= \frac{b-1}{b}\ a_{jj'}\ (\ -\frac{2}{b-1} + \left(\tfrac{1}{b-1}\right)^2 (b-2))\ \sigma^2$$

$$= \frac{-1}{b-1}\ a_{jj'}\ \sigma^2\ .$$

Substituting $a_{ij} = \frac{(b-1)}{bt}\ (t\ \delta_{ij} - 1)$ into the above results completes the derivation

$$E[z_{ij}\ z_{ij'}] = \frac{(b-1)(t\ \delta_{jj'} - 1)}{bt}\ \sigma^2$$

and

$$E[z_{ij}\ z_{i'j'}] = \frac{-(t\ \delta_{jj'} - 1)}{bt}\ \sigma^2$$

when $i \neq i'$. These two equations can be combined into the one equation

$$E[z_{ij}\ z_{i'j'}] = \frac{(b\ \delta_{ii'} - 1)(t\ \delta_{jj'} - 1)}{bt}\ \sigma^2 \ . \tag{5.13}$$

The identical formula holds for the infinite error model.

C. Third Order Moments

It was shown in the previous section that the finite error structure did not affect the variance-covariance structure of the residuals. The object of this section is to show that the same statement applies to the third order moments. In order to do this certain preliminary results are required. First

$$E[\delta_{ij}^{k}\ \delta_{ij'}^{k'}\ \delta_{ij''}^{k''}] \begin{cases} = \dfrac{1}{n} \text{ if } \ell=\ell'=\ell'' \text{ and } k=k'=k'' \\[2ex] = \dfrac{1}{n(n-1)} \text{ if } \ell=\ell'\neq\ell'' \text{ and } k=k'\neq k'' \\[2ex] = \dfrac{1}{n(n-1)(n-2)} \text{ if all } \ell \neq \text{ and all } k\neq \\[2ex] = 0 \ \textit{otherwise} \ . \end{cases}$$

The fact that the randomization is performed independently in each block ensures that $E[e_{ij}\ e_{i'j'}\ e_{i''j''}] = 0$ if $i\neq i'$, $i\neq i''$ or $i'\neq i''$.

Consequently only the term $E[e_{ij}\ e_{ij'}\ e_{ij''}]$ need be derived.

$$E[e_{ij}\ e_{ij'}\ e_{ij''}] = E[(\sum_k \delta_{ij}^k\ \varepsilon_{ik})(\sum_{k'} \delta_{ij'}^{k'}\ \varepsilon_{ik'})(\sum_{k''} \delta_{ij''}^{k''}\ \varepsilon_{ik''})] \tag{5.14}$$

$$= \sum_k E[\delta_{ij}^k\ \delta_{ij'}^k\ \delta_{ij''}^k]\ \varepsilon_{ik}^3$$

$$+ 3 \text{ terms like } \sum_{\substack{k,k' \\ k \neq k'}} E[\delta_{ij}^k\ \delta_{ij'}^k\ \delta_{ij''}^{k'}]\ \varepsilon_{ik}^2\ \varepsilon_{ik'}$$

$$+ \sum_{\substack{k,k'k'' \\ \text{all} \neq}} E[\delta_{ij}^k\ \delta_{ij'}^{k'}\ \delta_{ij''}^{k''}]\ \varepsilon_{ik}\ \varepsilon_{ik'}\ \varepsilon_{ik''}$$

$$= \delta_{jj'}\ \delta_{jj''}\ (\frac{1}{n}) \sum_k \varepsilon_{ik}^3$$

$$- [\delta_{jj'}\ (1-\delta_{jj''}) + \delta_{jj''}\ (1-\delta_{jj'}) + \delta_{jj''}\ (1-\delta_{jj'})]\ \frac{1}{n(n-1)} \sum_k \varepsilon_{ik}^3$$

$$+ 2[(1-\delta_{jj'})(1-\delta_{jj''})(1-\delta_{j'j''})]\ \frac{1}{n(n-1)(n-2)} \sum_k \varepsilon_{ik}^3$$

$$= [\delta_{jj'}\ \delta_{jj''} - \frac{1}{n}\ (\delta_{jj'} + \delta_{jj''}\ \delta_{j'j''}) + \frac{2}{n^2}]\ \frac{n}{(n-1)(n-2)} \sum_k \varepsilon_{ik}^3\ .$$

If it is assumed that the third order moments of the plot within block errors are equal and

$$\frac{n}{(n-1)(n-2)} \sum_k \varepsilon_{ik}^3 = M_3$$

then (5.14) can be written as

$$E[e_{ij}\ e_{ij'}\ e_{ij''}] \tag{5.15}$$

$$= [\delta_{jj'}\ \delta_{jj''} - \frac{1}{n}\ (\delta_{jj'} + \delta_{jj''} + \delta_{j'j''}) + \frac{2}{n^2}]\ M_3\ .$$

All possible third order moments can now be obtained by reasonably simple algebra and use of equation (5.15).

$$E[z_{ij}\ z_{ij'}\ z_{ij''}] \tag{5.16}$$

$$= E\Big[\Big(\sum_{\ell} a_{j\ell}\ [e_{i\ell} - \Big(\frac{1}{b-1}\Big) \sum_{h\neq i} e_{h\ell}]\Big)\Big(\sum_{\ell'} a_{j\ell'}\ [e_{i\ell'} - \Big(\frac{1}{b-1}\Big) \sum_{h\neq i} e_{h\ell'}]\Big)$$

$$\Big(\sum_{\ell''} a_{j\ell''}\ [e_{i\ell'} - \Big(\frac{1}{b-1}\Big) \sum_{h\neq i} e_{h\ell''}]\Big)\Big]$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell}\ \{E[e_{i\ell}^3] - \Big(\frac{1}{b-1}\Big)^3 \sum_{h\neq i} E[e_{h\ell}^3]\}$$

$$+\ 3 \text{ terms like} \sum_{\substack{\ell,\ell' \\ \ell\neq\ell'}} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell'}\ \{E[e_{i\ell}^2\ e_{i\ell'}] - \Big(\frac{1}{b-1}\Big)^3 \sum_{h\neq i} E[e_{h\ell}^2\ e_{h\ell'}]\}$$

$$+ \sum_{\substack{\ell,\ell',\ell'' \\ \text{all} \neq}} a_{j_\ell}\ a_{j'_{\ell'}}\ a_{j''_{\ell''}}\ \{E[e_{i\ell}\ e_{i\ell'}\ e_{i\ell''}]$$

$$- \Big(\frac{1}{b-1}\Big)^3 \sum_{h\neq i} E[e_{h\ell}\ e_{h\ell'}\ e_{h\ell''}]\}$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell}\ \frac{b(b-1)(b-2)}{(b-1)^3}\, M_3\ .$$

The case where two residuals are from one block and one from a second block leads to a similar equation. Here one obtains

$$\underset{i \neq i'}{E} [z_{ij}\ z_{ij'}\ z_{ij''}] \tag{5.17}$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell} \{ \cdot \left(\frac{-1}{b-1}\right) E[e^3_{i\ell}] + \left(\frac{1}{b-1}\right)^2 E[e^3_{i'\ell}]$$

$$- \left(\frac{1}{b-1}\right) \sum_{\substack{h \neq i \\ \neq i'}} E[e^3_{h\ell}]\}$$

$$+ 3 \text{ terms like } \sum_{\substack{\ell,\ell' \\ \ell \neq \ell'}} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell'} \{ \left(\frac{-1}{b-1}\right) E[e^2_{i\ell}\ e_{i\ell'}]$$

$$+ \left(\frac{1}{b-1}\right)^2 E[e^2_{i'\ell}\ e_{i'\ell'}] - \left(\frac{1}{b-1}\right)^3 \sum_{\substack{h \neq i \\ \neq i'}} E[e^2_{h\ell}\ e_{h\ell'}]\}$$

$$+ \sum_{\substack{\ell,\ell',\ell'' \\ \text{all} \neq}} a_{j\ell}\ a_{j'\ell'}\ a_{j''\ell''} \{ \left(\frac{-1}{b-1}\right) E[e_{i\ell}\ e_{i\ell'}\ e_{i\ell''}]$$

$$+ \left(\frac{1}{b-1}\right)^2 E[e_{i'\ell}\ e_{i'\ell'}\ e_{i'\ell''}] - \left(\frac{1}{b-1}\right)^3 \sum_{\substack{h \neq i \\ \neq i'}} E[e_{h\ell}\ e_{h\ell''}]\}$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell} \frac{-b(b-2)}{(b-1)^3} M_3 .$$

The case where all residuals are from different blocks follows the same pattern.

$$\underset{\text{all } i \neq}{E} [z_{ij}\ z_{i'j'}\ z_{i''j''}] = \sum_{\ell} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell} \frac{2b}{(b-1)^3} M_3 . \tag{5.18}$$

Equations (5.16), (5.17) and (5.18) can be combined into the one equation

$$E[z_{ij}\ z_{i'j'}\ z_{i''j''}] \tag{5.19}$$

$$= \sum_{\ell} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell}\ \{\delta_{ii'}\ \delta_{ii''}\ b(b-1)(b-2) - [\delta_{ii'}\ (1-\delta_{ii''})$$

$$+ \delta_{ii''}\ (1-\delta_{ii'}) + \delta_{i'i''}\ (1-\delta_{ii'})]\ b(b-2) + 2(1-\delta_{ii'})(1-\delta_{ii''})(1-\delta_{i'i''})b\}$$

$$\left(\frac{1}{b-1}\right)^3 M_3 .$$

However, from the definition of a_{ij} it follows that

$$\sum_{\ell} a_{j\ell}\ a_{j'\ell}\ a_{j''\ell} = \left(\frac{b-1}{bt}\right)^3 \sum_{\ell} (t\ \delta_{j\ell} - 1)(t\ \delta_{j'\ell} - 1)(t\ \delta_{j''\ell} - 1)$$

$$= \left(\frac{b-1}{bt}\right)^3 [t^3 \delta_{jj'}\ \delta_{jj''} - t^2(\delta_{jj'} + \delta_{jj''} + \delta_{j'j''}) + 2t] .$$

Collecting terms in (5.19) leads to the final result

$$E[z_{ij}\ z_{i'j'}\ z_{i''j''}] = \left(\frac{1}{bt}\right)^3 [t^3\ \delta_{jj'}\ \delta_{jj''} - t^2(\delta_{jj'} + \delta_{jj''} + \delta_{jj''}) + 2t]$$

$$[b^3\ \delta_{ii'}\ \delta_{ii''} - b^2(\delta_{ii'} + \delta_{ii''} + \delta_{i'i''}) + 2b]\ M_3 . \tag{5.20}$$

One of the more interesting features of this result is the fact that n does not appear. This implies that the third order moments are not influenced by restricting the number of possible errors. This holds even though the plot within block errors are correlated under this model and were independent under the former.

A summary of the expected values of the ten possible third order

Table 5. Summary of Expected Values of Third Order Moments

Moment	Restrictions on Subscripts	Expected Value
$E[z_{ij}^3]$	-	$\left(\frac{1}{bt}\right)^3 (t^3-3t^2+2t)(b^3-3b^2+2b)\ K_3$
$E[z_{ij}^2\ z_{ij'}]$	$j\neq j'$	$-\left(\frac{1}{bt}\right)^3 (t^2-2t)(b^3-3b^2+2b)\ K_3$
$E[z_{ij}\ z_{ij'}\ z_{ij''}]$	all $j\neq$	$\left(\frac{1}{bt}\right)^3 (2t)(b^2-3b^2+2b)\ K_3$
$E[z_{ij}^2\ z_{i'j}]$	$i\neq i'$	$-\left(\frac{1}{bt}\right)^3 (t^3-3t^2+2t)(b^2-2b)\ K_3$
$E[z_{ij}^2\ z_{i'j'}]$	$i\neq i', j\neq j'$	$\left(\frac{1}{bt}\right)^3 (t^2-2t)(b^2-2b)\ K_3$
$E[z_{ij}\ z_{ij'}\ z_{i'j}]$	$i\neq i', j\neq j'$	$\left(\frac{1}{bt}\right)^3 (t^2-2t)(b^2-2b)\ K_3$
$E[z_{ij}\ z_{ij'}\ z_{i'j''}]$	$i\neq i'$, all $j\neq$	$-\left(\frac{1}{bt}\right)^3 (2t)(b^2-2b)\ K_3$
$E[z_{ij}\ z_{i'j}\ z_{i''j}]$	all $i\neq$	$\left(\frac{1}{bt}\right)^3 (t^3-3t^2+2t)(2b)\ K_3$
$E[z_{ij}\ z_{i'j}\ z_{i''j'}]$	all $i\neq, j\neq j'$	$-\left(\frac{1}{bt}\right)^3 (t^2-2^t)(2b)\ K_3$
$e[z_{ij}\ z_{i'j'}\ z_{i''j''}]$	all $i\neq$, all $j\neq$	$\left(\frac{2}{bt}\right)^2\ K_3$

moments possible among the residuals in a randomized block design is given in Table 5. These moments can be contrasted with the expected values of these moments under the assumption of independent normal errors. The moments all have zero expectations under this alternate model.

D. Fourth Order Moments

The previous two sections demonstrated that the second and third order moments do not depend on n, the number of plots per block. This, however, does not hold for the fourth order moments. Since there is a total of 25 different fourth order moments, only one, the expected value of the fourth power of the residuals will be obtained. The method employed is similar to that used in sections B and C.

The first step is to obtain the moments of the btn random variables $\{\delta_{ij}^{k}\}$.

$$E[\delta_{i\ell_1}^{k_1}\delta_{i\ell_2}^{k_2}\delta_{i\ell_3}^{k_3}\delta_{i\ell_4}^{k_4}] \begin{cases} = \frac{1}{n} & \text{if } k_1=k_2=k_3=k_4,\ \ell_1=\ell_2=\ell_3=\ell_4 \\ = \frac{1}{n(n-1)} & \text{if } k_1=k_2=k_3\neq k_4,\ \ell_1=\ell_2=\ell_3\neq\ell_4 \\ = \frac{1}{n(n-1)} & \text{if } k_1=k_2\neq k_3=k_4,\ \ell_1=\ell_2\neq\ell_3=\ell_4 \\ = \frac{1}{n(n-1)(n-2)} & \text{if } k_1=k_2,\ \ell_1=\ell_2 \text{ and all other k's and } \ell\text{'s are unequal} \\ = \frac{1}{n(n-1)(n-2)(n-3)} & \text{if all k's and all } \ell\text{'s are unequal} \\ = 0 & \text{otherwise} \end{cases} \tag{5.21}$$

Also $E[\delta_{i_1\ell_1}^{k_1} \delta_{i_1\ell_2}^{k_2} \delta_{i_2\ell_3}^{k_3} \delta_{i_2\ell_4}^{k_4}]$ (5.22)

$$= E[\delta_{i_1\ell_1}^{k_1} \delta_{i_1\ell_2}^{k_2}]\; E[\delta_{i_2\ell_3}^{k_3} \delta_{i_2\ell_4}^{k_4}] \text{ if } i_1 \neq i_2 \;.$$

Finally expressions in which one of the i subscripts differs from all others can be ignored since the randomization and hence the errors are independent in different blocks.

In the following derivation it will be assumed that j_1, j_2, j_3, and j_4 are all unequal. It will also be assumed that the second and fourth moments of the plot within block errors are similar enough that they can be treated as equal. In particular let

$$S_2 = \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{ik}^2$$

and

$$S_4 = \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{ik}^4 \;.$$

In general

$$E[e_{ij_1}\, e_{ij_2}\, e_{ij_3}\, e_{ij_4}] = \tag{5.23}$$

$$E[(\sum_{k_1} \delta_{ij_1}^{k_1}\, \varepsilon_{ik_1})(\sum_{k_2} \delta_{ij_2}^{k_2}\, \varepsilon_{ik_2})(\sum_{k_3} \delta_{ij_3}^{k_3}\, \varepsilon_{ik_3})(\sum_{k_4} \delta_{ij_4}^{k_4}\, \varepsilon_{ik_4})] \;.$$

From this the special cases follow:

$$E[e_{ij}^4] = S_4 \;. \tag{5.24}$$

$$E[e_{ij_1}^3\, e_{ij_2}] = -\frac{1}{n-1} S_4 \;. \tag{5.25}$$

$$E[e_{ij_1}^2 \; e_{ij_2}^2] = \frac{1}{n-1} [n \; S_2^2 - S_4] \tag{5.26}$$

$$E[e_{ij_1}^2 \; e_{ij_2} \; e_{ij_3}] = \frac{1}{(n-1)(n-2)} [2 \; S_4 - n \; S_2^2] \tag{5.27}$$

$$E[e_{ij_1} \; e_{ij_2} \; e_{ij_3} \; e_{ij_4}] = \frac{3}{(n-1)(n-2)(n-3)} [n \; S_2^2 - 2 \; S_4] \tag{5.28}$$

If $i_1 \neq i_2$ then

$$E[e_{i_1j_1} \; e_{i_1j_2} \; e_{i_2j_3} \; e_{i_2j_4}] = E[e_{i_1j_1} \; e_{i_1j_2}] \; E[e_{i_2j_3} \; e_{i_2j_4}] \tag{5.29}$$

$$= E[(\sum_{k_1} \delta_{i_1j_1}^{k_1} \; \varepsilon_{i_1k_1})(\sum_{k_2} \delta_{i_1j_2}^{k_2} \; \varepsilon_{i_1k_2})] \; E[(\sum_{k_3} \delta_{i_2j_3}^{k_3} \; \varepsilon_{i_2k_3})(\sum_{k_4} \delta_{i_2j_1}^{k_4} \; \varepsilon_{i_2k_4})].$$

It follows that

$$E[e_{i_1j}^2 \; e_{i_2j}^2] = S_2^2 \; , \tag{5.30}$$

$$E[e_{i_1j}^2 \; e_{i_2j_1} \; e_{i_2j_2}] = - \frac{1}{n-1} \; S_2^2 \; , \tag{5.31}$$

and

$$E[e_{i_1j_1} \; e_{i_1j_2} \; e_{i_2j_1'} \; e_{i_2j_2'}] = \left(\frac{1}{n-1}\right)^2 \; S_2^2 \; . \tag{5.32}$$

In equation (5.32) the restriction $j_1' \neq j_2'$ also applies.

Now write

$$E[z_{ij}^4] = E[(\sum_{\ell} a_{j\ell}[e_{i\ell} - \frac{1}{b-1} \sum_{h \neq i} e_{h\ell}])^4] \tag{5.33}$$

$$= \sum_{\ell} a_{j\ell} \left\{ E[e_{i\ell}^4] + \left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} E[e_{h\ell}^4] + 3\left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} \sum_{\substack{h\neq i \\ \neq h}} E[e_{h\ell}^2]\, E[e_{h'\ell}^2] \right.$$

$$\left. + 6\left(\frac{1}{b-1}\right)^2 E[e_{i\ell}^2] \sum_{h\neq i} E[e_{h\ell}^2] \right\}$$

$$+ 4 \sum_{\substack{\ell_1,\ell_2 \\ \ell_1\neq\ell_2}} a_{j\ell_1}^3 \left\{ E[e_{i\ell_1}^3\, e_{i\ell_2}] + \left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} E[e_{h\ell_1}^3\, e_{h\ell_2}] \right.$$

$$+ 3\left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} \sum_{\substack{h'\neq i \\ \neq h}} E[e_{h\ell_1}^2]\, E[e_{h'\ell_1}\, e_{h'\ell_2}]$$

$$+ 3\left(\frac{1}{b-1}\right)^2 E[e_{i\ell_1}^2] \sum_{h\neq i} E[e_{h\ell_1}\, e_{h\ell_2}]$$

$$\left. + 3\left(\frac{1}{b-1}\right)^2 E[e_{i\ell_1}\, e_{i\ell_2}] \sum_{h\neq 1} E[e_{h\ell_1}^2] \right\}$$

$$+ 3 \sum_{\substack{\ell_1,\ell_2 \\ \ell_1\neq\ell_2}} a_{j\ell_1}^2\, a_{j\ell_2}^2 \left\{ E[e_{i\ell_1}^2\, e_{i\ell_2}^2] + \left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} E[e_{h\ell_1}^2\, e_{h\ell_2}^2] \right.$$

$$+ \left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} \sum_{\substack{h\neq i \\ \neq h}} E[e_{h\ell_1}^2]\, E[e_{h'\ell_2}^2]$$

$$+ 2\left(\frac{1}{b-1}\right)^4 \sum_{h\neq 1} \sum_{\substack{h\neq 1 \\ \neq h}} E[e_{h\ell_1}\, e_{h\ell_2}]\, E[e_{h'\ell_1}\, e_{h'\ell_2}]$$

$$+ 2\left(\frac{1}{b-1}\right)^2 E[e_{i\ell_1}^2] \sum_{h\neq i} E[e_{h\ell_2}^2]$$

$$\left. + 4\left(\frac{1}{b-1}\right)^2 E[e_{i\ell_1}\, e_{i\ell_2}] \sum_{h\neq i} E[e_{h\ell_1}\, e_{h\ell_2}] \right\}$$

$$+ 6 \sum_{\substack{\ell_1,\ell_2,\ell_3 \\ \text{all} \neq}} a_{j\ell_1}^2 \; a_{j\ell_2} \; a_{j\ell_3} \Big\{ E[e_{i\ell_1}^2 \; e_{i\ell_2} \; e_{i\ell_3}]$$

$$+ \left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} E[e_{h\ell_1} \; e_{h\ell_2} \; e_{h\ell_3}]$$

$$+ \left(\frac{1}{b-1}\right)^2 E[e_{i\ell_1}^2] \sum_{h=i} E[e_{h\ell_2} \; e_{h\ell_3}]$$

$$+ \left(\frac{1}{b-1}\right)^2 E[e_{i\ell_2} \; e_{i\ell_3}] \sum_{h\neq i} E[e_{h\ell_1}^2]$$

$$+ 4\left(\frac{1}{b-1}\right)^2 E[e_{i\ell_1} \; e_{i\ell_2}] \sum_{h\neq i} E[e_{h\ell_1} \; e_{h\ell_3}]$$

$$+ \left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} \sum_{\substack{h'\neq i \\ \neq h}} E]e_{h\ell_1}^2] \; E[e_{h'\ell_2} \; e_{h'\ell_3}]$$

$$+ 2\left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} \sum_{\substack{h'\neq i \\ \neq h}} E[e_{h\ell_1} \; e_{h\ell_2}] \; E[e_{h'\ell_1} \; e_{h'\ell_3}]\Big\}$$

$$+ \sum_{\substack{\ell_1,\ell_2,\ell_3,\ell_4 \\ \text{all} \neq}} a_{j\ell_1} \; a_{j\ell_2} \; a_{j\ell_3} \; a_{j\ell_4} \Big\{ E[e_{i\ell_1} \; e_{i\ell_2} \; e_{i\ell_3} \; e_{i\ell_4}]$$

$$+ \left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} E[e_{h\ell_1} \; e_{h\ell_2} \; e_{h\ell_3} \; e_{h\ell_4}]$$

$$+ 6\left(\frac{1}{b-1}\right)^2 E[e_{i\ell_1} \; e_{i\ell_2}] \sum_{h\neq i} E[e_{h\ell_3} \; e_{h\ell_4}]$$

$$+ 3\left(\frac{1}{b-1}\right)^4 \sum_{h\neq i} \sum_{\substack{h'\neq i \\ \neq h}} E[e_{h\ell_1} \; e_{h\ell_2}] \; E[e_{h\ell_3} \; e_{h\ell_4}]\Big\}$$

$$= \sum_{\ell} a_{j\ell}^4 \left\{ [1 + (\tfrac{1}{b-1})^3]\ S_4 + \left[\frac{3b(2b-3)}{(b-1)^3}\right] S_2^2 \right\}$$

$$+ 4 \sum_{\substack{\ell_1,\ell_2 \\ \ell_1 \neq \ell_2}} a_{j\ell_1}^3\ a_{j\ell_2} \left\{ - \frac{1}{(n-1)}\ [1 + (\tfrac{1}{b-1})^3]\ S_4 - \frac{1}{(n-1)} \left[\frac{3b(2b-3)}{(b-1)^3}\right] S_2^2 \right\}$$

$$+ 3 \sum_{\substack{\ell_1,\ell_2 \\ \ell_1 \neq \ell_2}} a_{j\ell_1}^2\ a_{j\ell_2}^2 \left\{ \frac{-1}{(n-1)}\ [1 + (\tfrac{1}{b-1})^3][S_4 - n\ S_2^2] + \frac{b(2b-3)}{(b-1)^3}\ S_2^2 \right.$$

$$\left. + (\tfrac{1}{n-1})^2 \left[\frac{2b(2b-3)}{(b-1)^3}\right] S_2^2 \right\}$$

$$+ 6 \sum_{\substack{\ell_1,\ell_2,\ell_3 \\ \text{all} \neq}} a_{j\ell_1}^2\ a_{j\ell_2}\ a_{j\ell_3} \left\{ \frac{1}{(n-1)(n-2)}\ [1 + (\tfrac{1}{b-1})^3][2\ S_4 - n\ S_2^2] \right.$$

$$\left. - (\tfrac{1}{n-1}) \left[\frac{b(2b-3)}{(b-1)^3}\right] S_2^2 + (\tfrac{1}{n-1})^2 \left[\frac{2b(2b-3)}{(b-1)^3}\right] S_2^2 \right\}$$

$$+ \sum_{\substack{\ell_1,\ell_2,\ell_3,\ell_4 \\ \text{all} \neq}} a_{j\ell_1}\ a_{j\ell_2}\ a_{j\ell_3}\ a_{j\ell_4} \left\{ \frac{-3}{(n-1)(n-2)(n-3)}\ [1 + (\tfrac{1}{b-1})^3][2\ S_4 - n\ S_2^2] \right.$$

$$\left. + (\tfrac{1}{n-1})^2 \left[\frac{3b(2b-3)}{(b-1)^3}\right] S_2^2 \right\}$$

$$= \sum_{\ell} a_{j\ell}^4 \left\{ \left[\frac{n^2(n+1)}{(n-1)(n-2)(n-3)}\right] [1 + (\tfrac{1}{b-1})^3]\ S_4 \right.$$

$$\left. - \left[\frac{3n^2}{(n-2)(n-3)}\right] [1 + (\tfrac{1}{b-1})^3]\ S_2^2 \right\}$$

$$+ (\sum_{\ell} a_{j\ell}^2)^2 \left\{ \left[\frac{-3n}{(n-2)(n-3)}\right] [1 + (\tfrac{1}{b-1})^3]\ S_4 \right.$$

$$+ 3 \left(\frac{n}{n-1}\right) [1 + \frac{2n-3}{(n-2)(n-3)}][1 + \left(\frac{1}{b-1}\right)^3] S_2^2$$

$$+ 3 \left(\frac{n}{n-1}\right)^2 \left[\frac{b(2b-3)}{(b-1)^3}\right] S_2^2 \Bigg\} .$$

Now define

$$K_4^* = \frac{n^2(n+1)}{(n-1)(n-2)(n-3)} S_4 - \frac{3n^2}{(n-2)(n-3)} S_2^2 .$$

Also recall that $S_2 = \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{ik}^2 = \frac{n-1}{n} \sigma^2$. With these definitions (5.33) can be rewritten as

$$E[z_{ij}^4] = \sum_{\ell} a_{j\ell}^4 [1 + \left(\frac{1}{b-1}\right)^3] K_4^* \tag{5.34}$$

$$+ \left(\sum_{\ell} a_{j\ell}^2\right)^2 \Bigg\{ \left[\frac{-3n}{(n-2)(n-3)}\right] [1 + \left(\frac{1}{b-1}\right)^3] S_4$$

$$+ 3 \left(\frac{n-1}{n}\right) [1 + \frac{2n-3}{(n-2)(n-3)} \quad [1 + \left(\frac{1}{b-1}\right)^3] \sigma^4$$

$$+ 3 \left[\frac{b(2b-3)}{(b-1)^3}\right] \sigma^4 \Bigg\}$$

$$= \sum_{\ell} a_{j\ell}^4 [1 + \left(\frac{1}{b-1}\right)^3] K_4^*$$

$$+ \left(\sum_{\ell} a_{j\ell}^2\right)^2 \Bigg\{ \left[\frac{-3n}{(n-2)(n-3)}\right] [1 + \left(\frac{1}{b-1}\right)^3] S_4$$

$$+ 3 \frac{n^2-3}{n(n-2)(n-3)} [1 + \left(\frac{1}{b-1}\right)^3] \sigma^4$$

$$+ 3 \left(\frac{b}{b-1}\right)^2 \sigma^4 \Bigg\} .$$

The fact that $a_{ij} = \frac{(b-1)}{bt} (t\ \delta_{ij} - 1)$ permits (5.34) to be written as

$$E[z_{ij}^4] = \frac{(b-1)(t-1)}{b^4t^4} [1 + (b-1)^3][1 + (t-1)^3]\ K_4^* \tag{5.35}$$

$$+ 3\ \frac{(b-1)(t-1)^2}{b^4t^2} [1 + (b-1)^3] \left\{ \left[\frac{n^2-3}{n(n-2)(n-3)}\right]\sigma^4 - \left[\frac{n}{(n-2)(n-3)}\right] S_4 \right\}$$

$$+ 3\ \frac{(b-1)^2(t-1)^2}{b^2t^2}\ \sigma^4 .$$

As n becomes large K_4^* approaches the fourth cumulant of the error distribution and $E[z_{ij}^4]$ approaches

$$\frac{(b-1)(t-1)}{b^4t^4} [1 + (b-1)^3][1 + (t-1)^3]\ K_4$$

$$+ 3\ \frac{(b-1)^2(t-1)^2}{b^2t^2}\ \sigma^4 .$$

This agrees with the expected value of the fourth power of the residuals obtained by Anscombe (1961). The second term in the expression on the right hand side of (5.35) is of interest. Clearly, this term will go to the zero as n becomes large. Except for the constant term which is always positive and can be written as

$$\frac{n}{(n-2)(n-3)} \left\{ \frac{n^2-3}{n^2} \sigma^4 - S_4 \right\}$$

$$= \frac{n}{(n-2)(n-3)} \left\{ (1 - \frac{3}{n^2}) \left(\frac{n-1}{n}\right)^2 S_2^2 - S_4 \right\}$$

$$= \frac{n}{(n-2)(n-3)} \left\{ S_2^2 - S_4 - (\frac{2}{n} + \frac{2}{n^2} - \frac{6}{n^3} + \frac{3}{n^4}) S_2^2 \right\} .$$

From the definitions of S_2 and S_4 it follows that $S_4 \geq S_2^2$. Also the term $(\frac{2}{n} + \frac{2}{n^2} - \frac{6}{n^3} + \frac{3}{n^4}) > 0$ for integral values of $n \geq 1$.

Consequently the whole expression will always be negative for $n > 3$. It follows that the fourth moment of the residuals is reduced slightly by this term under the finite error model.

The variance of the square of the residuals can be obtained immediately from (5.35) by using the fact that

$$E[z_{ij}^2] = \frac{(b-1)(t-1)}{bt} \sigma^2 .$$

VI. TESTS FOR NON-ADDITIVITY

A. Introduction

Properties of the residuals under two competing models were examined in the previous chapters. In this chapter some tests for non-additivity in the r x c table with one observation per cell will be developed. The approach is to partition the error sum of squares into two independent terms in such a manner that one of them will be inflated relative to the other if non-additivity is present.

Before proceeding with the development of the various tests for non-additivity, certain preliminary results regarding the partitioning of the total sum of squares are required. In matrix notation one can write a linear model as

$$y = X\,\beta + e \tag{6.1}$$

where y is the n x 1 vector of observations, X is the n x p matrix of known constants, β is a p x 1 vector of unknown parameters and e is an n x 1 vector of unknown errors. It will be assumed that the matrix X is of rank p. If the rank of X is less than p, the formula $(X'X)^{-1}$ will denote a conditional or generalized inverse of X'X and p will be used for its rank.

Under the assumption that the elements of the vector e are independent with mean zero and equal variance, the Gauss-Markoff theorem states that the best linear unbiased estimate of the elements of the vector X β are given by the equation

$$\widehat{X\,\beta} = X(X'X)^{-1}\,X'y \tag{6.2}$$

Since the rank of X was assumed to be p , $(p < n)$ it follows that the vector $X\hat{\beta}$ spans a p-dimensional subspace of the n-dimensional space spanned by the original observations. Consequently, the practice of estimating the elements of $X \beta$ is really a procedure for partitioning the space spanned by the observations into two parts. One of these is referred to as the estimation space and is spanned by $X(X'X)^{-1} X'y$ and the other the error space, spanned by

$$y - X(X'X)^{-1} X'y = [I - X(X'X)^{-1} X'] y . \qquad (6.3)$$

It follows from the equation

$$X(X'X)^{-1}X' [I - X(X'X)^{-1} X'] = 0 \qquad (6.4)$$

that the two spaces are orthogonal. Further, if it is assumed that e is multivariate normal then tests of hypotheses about the parameters can be constructed by comparing functions from the parameter space with functions of the residuals in the error space.

If the model specified in (6.1) is in fact correct, i.e., additivity holds, then the error space will be perfectly free of the effects of the parameters. If, however, this is not true then one should be able to use information about the parameters, i.e., their estimates to compute a contrast among the residuals which would account for a disproportionately large fraction of the error sum of squares. Under the assumptions of additivity and normal independent errors, the conditional distribution of the ratio of the sum of squares due to any special contrast selected on the basis of the parameter estimates and the remaining portion of the error sum of squares is that of a constant times an F variate with

1 and n - p - 1 degrees of freedom. However, the unconditional distribution is the same F distribution, because the conditional distribution is functionally independent of the conditioning variables.

Let

$$z = [I - X(X'X)^{-1} X'] y \tag{6.5}$$

where x and y are defined as in (6.1) and I is the n X n identity matrix. Also assume that

$$z = H \theta + e^{*} \tag{6.6}$$

where H is a known n x q matrix of rank q, which may or may not depend on $\hat{\delta}$, θ is a q x 1 vector of unknown parameters, e^{*} is an n x 1 multivariate normal vector with mean zero and variance-covariance matrix

$$Q \sigma^2 = [I - X(X'X)^{-1} X'] \sigma^2 .$$

If S is the sum of squares due to fitting θ in the model (6.6), then S/σ^2 has, conditionally on $X \hat{\beta}$, a chi-square distribution with q degrees of freedom under the hypothesis that $\theta = 0$ if $H'X(X'X)^{-1}X' = 0$ (a null matrix). This is seen by noting that $z'H(H'H)^{-1}H'z$ equals $y'Q'H(H'H)^{-1}H'Qy$. Theorem 4.7 of Graybill states that $y'Ay$ is distributed as X^2 if A is idempotent when V(y) is σ^2 I. The matrix $Q'H(H'H)^{-1}H' Q$ is indeed idempotent because

$Q'H(H'H)^{-1} H'Q = [I - X(X'X)^{-1} X'] H(H'H)^{-1} H' [I - X(X'X)^{-1} X']$
which equals $H(H'H)^{-1} H'$ under the condition $H'X(X'X)^{-1} = 0$. Also $H(H'H)^{-1} H'$ is clearly idempotent.

Various tests for non-additivity can be obtained as special cases of the above theorem. A common test due to Tukey (1949), Moore and Tukey (1954) and Tukey (1955) is an example. This test and a number of modifications obtained by making different choices for the matrix H and the vector θ will be discussed in the remainder of this chapter.

B. Tukey's Test for Non-additivity

Tukey's test for non-additivity can be motivated by assuming that a non-linear function of the additive $\{y_i\}$ is observed. It is assumed further that this function can be represented adequately by

$$h(y_i) = y_i + \phi(y_i - \mu_i)^2 \tag{6.7}$$

where μ_i denotes the expected value of y_i for i = 1, 2, ..., n . Under these conditions it is shown by Anscombe (1961) that the statistic

$$f = \frac{\sum_i z_i Y_i^2}{\sum_{ij} q_{ij} Y_i^2 Y_j^2} \tag{6.8}$$

where the $\{Y_i\}$ are the elements of the vector $X(X'X)^{-1} X'y$ and the $\{q_{ij}\}$ are the elements of the matrix $I - X(X'X)^{-1} X'$, is a rough estimate of ϕ . It is also shown that

$$\frac{(\sum_i z_i Y_i^2)^2}{\sum_{ij} q_{ij} Y_i^2 Y_j^2}$$

is a one-degree-of-freedom component which can be removed from the residual sum of squares, leaving an independent remainder of $n - p - 1$ degrees of freedom. The ratio of these two components constitutes Tukey's test for non-additivity.

When the data is obtained from an $r \times c$ table with one observation per cell, the formula to compute the sum of squares due to non-additivity simplifies considerably. Let

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \tag{6.9}$$

be the model for the observations, where the $\{\alpha_i\}$ and $\{\beta_j\}$ are the deviations due to row and column effects respectively and μ the over-all mean. Now define

$$z_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..} \tag{6.10}$$

for all $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$. The sum of squares due to non-additivity for Tukey's test can then be written in the more familiar form

$$\frac{[\sum_{ij} z_{ij}(y_{i.} - y_{..})(y_{.j} - y_{..})]^2}{[\sum_i (y_{i.} - y_{..})^2][\sum_j (y_{.j} - y_{..})^2]} . \tag{6.11}$$

Scheffé (1959) uses a slightly different approach to motivate Tukey's test for the two-way table. He begins with the model

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ij} .$$

Ideally he would like to test whether all $\gamma_{ij} = 0$. This, however, is not possible. Consequently, he restricts the interaction term to be of the form

$$\gamma_{ij} = G\, \alpha_i\, \beta_j \ ,$$

where G is a constant.

The complete model then becomes

$$y_{ij} = \mu + \alpha_i + \beta_j + G\, \alpha_i\, \beta_j + e_{ij} \ .$$

At this point one might be tempted to use a least squares procedure to estimate all parameters simultaneously. While this would lead to satisfactory estimates of the parameters, it does not lead to any simple tests of hypothesis. The approach used by Scheffé is to estimate μ , $\{\alpha_i\}$ and $\{\beta_j\}$ by least squares, pretend that these estimates are the real parameters and proceed to estimate G . A test for G , which is really Tukey's test is then obtained, first conditionally and then unconditionally on the initial set of parameter estimates.

These two arguments leading to the same test are both given here, since both types of reasoning will be used to motivate some of the tests to be developed.

C. Other Tests for Non-additivity

The remainder of this chapter will be restricted to the $r \times c$ classification with one observation per cell. If non-additivity is present, the most general model is

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ij} \tag{6.12}$$

where $\alpha_. = \beta_. = \gamma_{i.} = \gamma_{.j} = 0$.

The notation used here is that if a dot replaces a subscript then an average has been taken over that subscript. This model, however, is of little interest since it is not possible to use it to test for possible non-additivity. However, an approach along the lines followed by Scheffé is possible. Assume that γ_{ij} is a function of α_i and β_j . The model can then be written as

$$y_{ij} = \mu + \alpha_i + \beta_j + f(\alpha_i\ , \beta_j) + e_{ij}\ . \tag{6.13}$$

Assume further that the function f can be replaced by a polynomial of degree m . This means that

$$\gamma_{ij} = \sum_{h=0}^{m} \sum_{h'=0}^{m-h} A_{hh'}\, \alpha_i^h\, \beta_j^{h'} \tag{6.14}$$

for all i = 1, 2, ..., r and j = 1, 2, ..., c . However, the linear dependencies among the $\{\gamma_{ij}\}$ give rise to some restrictions on the $\{A_{hh'}\}$. If α^h is used to represent $\frac{1}{r}\sum_i \alpha_i^h$ and β^h to represent $\frac{1}{c}\sum_j \beta_j^h$ then one can write

$$\gamma_{ij} = \gamma_{ij} - \gamma_{i.} - \gamma_{.j} + \gamma_{..} \tag{6.15}$$

$$= \sum_{h=1}^{m-1} \sum_{h'=1}^{m-h} A_{hh'}(\alpha_i^h - \alpha^h)(\beta_j^{h'} - \beta^{h'})\ .$$

Of course, $\alpha^1 = \beta^1 = 0$.

The model can now be written as

$$y_{ij} = \mu + \alpha_i + \beta_j + \sum_{h=1}^{m-1} \sum_{h'=1}^{m-h} A_{hh'}(\alpha_i^h - \alpha^h)(\beta_j^{h'} - \beta^{h'}) + e_{ij} \ . \quad (6.16)$$

This model immediately suggests a series of tests for non-additivity, motivated along the lines used by Scheffé (1959). Tukey's test is obtained by taking $m = 2$. For $m = 3$ (6.16) becomes

$$y_{ij} = \mu + \alpha_i + \beta_j + A_{11}\,\alpha_i\,\beta_j + A_{12}\,\alpha_i(\beta_j^2 - \beta^2)$$

$$+ A_{21}(\alpha_i^2 - \alpha^2)\beta_j + e_{ij} \ . \quad (6.17)$$

Estimates for α_i and β_j are obtained as the deviations of the appropriate row and column means from the grand mean. The residuals can be written as

$$z_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..} \quad (6.18)$$

$$= A_{11}\,\alpha_i\,\beta_j + A_{12}\,\alpha_i(\beta_j^2 - \beta^2) + A_{21}(\alpha_i^2 - \alpha^2)\beta_j$$

$$+ e_{ij} - e_{i.} - e_{.j} + e_{..} \ .$$

If the $\{\alpha_i\}$ and $\{\beta_j\}$ in (6.18) are replaced by their estimates, the theorem proved in section A of this chapter can be applied, to obtain a test for non-additivity with 3 degrees of freedom. The sum of squares due to non-additivity is equal to

$$N = \hat{A}_{11}(\sum z_{ij}\ \hat{\alpha}_i\ \hat{\beta}_j) + \hat{A}_{12}(\sum z_{ij}\ \hat{\alpha}_i(\hat{\beta}_j^2 - \hat{\beta}^2))$$

$$+ \hat{A}_{21}(\sum z_{ij}(\hat{\alpha}_i^2 - \alpha^2)\hat{\beta}_j) \tag{6.19}$$

where $\hat{A}_{11}$, $\hat{A}_{12}$ and $\hat{A}_{21}$ are the solutions to the least squares equation

$$\begin{pmatrix} \sum \hat{\alpha}_i^2\ \hat{\beta}_j^2 & \sum \hat{\alpha}_i^2\ \hat{\beta}_j(\hat{\beta}_j^2 - \hat{\beta}^2) & \sum \hat{\alpha}_i(\hat{\alpha}_i^2 - \hat{\alpha}^2)\beta_j^2 \\ \sum \hat{\alpha}_i^2\ \hat{\beta}_j^2(\hat{\beta}_j^2 - \hat{\beta}^2) & \sum \hat{\alpha}_i^2(\hat{\beta}_j^2 - \hat{\beta}^2)^2 & \sum \hat{\alpha}_i\ \hat{\beta}_j(\hat{\alpha}_i^2 - \hat{\alpha}^2)(\hat{\beta}_j^2 - \hat{\beta}^2) \\ \sum \hat{\alpha}_i(\hat{\alpha}_i^2 - \hat{\alpha}^2)\hat{\beta}_j^2 & \sum \hat{\alpha}_i\ \hat{\beta}_j(\hat{\alpha}_i^2 - \hat{\alpha}^2)(\hat{\beta}_j^2 - \hat{\beta}^2) & \sum(\hat{\alpha}_i^2 - \hat{\alpha}^2)^2\ \hat{\beta}_j^2 \end{pmatrix} \begin{pmatrix} A_{11} \\ A_{12} \\ A_{21} \end{pmatrix} =$$

$$\begin{pmatrix} \sum z_{ij}\ \hat{\alpha}_i\ \hat{\beta}_j \\ \sum z_{ij}\ \hat{\alpha}_i(\hat{\beta}_j^2 - \hat{\beta}^2) \\ \sum z_{ij}(\hat{\alpha}_i^2 - \hat{\alpha}^2)\hat{\beta}_j \end{pmatrix} . \tag{6.20}$$

If $S = \sum z_{ij}^2$ then the statistic

$$\frac{(r-1)(c-1) - 3}{3} \times \frac{N}{S-N}$$

has an F distribution with 3 and $(r-1)(c-1) - 3$ degrees of freedom under the null hypothesis that $A_{11} = A_{12} = A_{21} = 0$.

A similar test for non-additivity can be derived if it is assumed that the effects are additive, i.e.

$$\mu_{ij} = \mu + \alpha_i + \beta_j \ , \tag{6.21}$$

but that the experimenter observes

$$y_{ij} = f(\mu_{ij}) + e_{ij} \ , \tag{6.22}$$

where f can be approximated by the series expansion

$$y_{ij} = \mu_{ij} + \phi\mu_{ij}^2 + \delta\mu_{ij}^3 + e_{ij} \ . \tag{6.23}$$

This model can be rewritten as

$$\begin{aligned} y_{ij} = {} & [\mu + \phi\mu^2 + \delta\mu^3 + (\phi + 3\delta\mu)(\alpha^2 + \beta^2) + \delta(\alpha^3 + \beta^3)] \\ & + [(1 + 2\phi\mu + 3\delta\mu^2)\alpha_i + (\phi + 3\delta\mu)(\alpha_i^2 - \alpha^2) + \delta(\alpha_i^3 - \alpha^3)] \\ & + [(1 + 2\phi\mu + 3\delta\mu^2)\beta_j + (\phi + 3\delta\mu)(\beta_j^2 - \beta^2) + \delta(\beta_j^3 - \beta^3)] \\ & + [(2\phi + 6\delta\mu)\alpha_i\,\beta_j + 3\delta\alpha_i(\beta_j^2 - \beta^2) + 3\delta(\alpha_i^2 - \alpha^2)\beta_j] \ . \end{aligned} \tag{6.24}$$

If the terms $(\phi + 3\delta\mu)$ and δ are small relative to $(1 + 2\phi\mu + 3\delta\mu^2)$, then (6.24) can be approximated by the model

$$\begin{aligned} y_{ij} = {} & \mu^* + A_{10}\,\alpha_i + A_{01}\,\beta_j + A_{11}\,\alpha_i\,\beta_j + A_{12}\,\alpha_i(\beta_j^2 - \beta^2) \\ & + A_{21}(\alpha_i^2 - \alpha^2)\beta_j + e_{ij} \end{aligned}$$

where the A_{hk} are constants. This model leads to the same test for non-additivity as the one spelled out in (6.17).

A slightly different series of tests for non-additivity can be obtained by simply defining the sum of squares due to non-additivity as the portion of the total variation in the residuals accounted for by regression on a series of terms like $\hat{\alpha}_i^h \hat{\beta}_j^k$ for all i and j . For example, a test for non-additivity with 3 degrees of freedom is obtained by finding the portion of the variation in the residuals accounted for by

$\hat{\alpha}_i \hat{\beta}_j$, $\hat{\alpha}_i^2 \hat{\beta}_j$ and $\hat{\alpha}_i \hat{\beta}_j^2$ simultaneously. The test statistic is

$$\frac{(r-1)(c-1) - 3}{3} \times \frac{N^*}{S - N^*}$$

where $S = \sum z_{ij}^2$,

$$N^* = \hat{B}_{11}(\sum z_{ij} \hat{\alpha}_i \hat{\beta}_j) + \hat{B}_{21}(\sum z_{ij} \hat{\alpha}_i^2 \hat{\beta}_j) + \hat{B}_{12}(\sum z_{ij} \hat{\alpha}_i \hat{\beta}_j^2) \qquad (6.25)$$

and $\hat{B}_{11}$, $\hat{B}_{21}$ and $\hat{B}_{12}$ are solutions to the equation

$$\begin{pmatrix} \sum \hat{\alpha}_i^2 \hat{\beta}_j^2 & \sum \hat{\alpha}_i^3 \hat{\beta}_j^2 & \sum \hat{\alpha}_i^2 \hat{\beta}_j^3 \\ \sum \hat{\alpha}_i^3 \hat{\beta}_j^2 & \sum \hat{\alpha}_i^4 \hat{\beta}_j^2 & \sum \hat{\alpha}_i^3 \hat{\beta}_j^3 \\ \sum \hat{\alpha}_i^2 \hat{\beta}_j^3 & \sum \hat{\alpha}_i^3 \hat{\beta}_j^3 & \sum \hat{\alpha}_i^2 \hat{\beta}_j^4 \end{pmatrix} \begin{pmatrix} \hat{B}_{11} \\ \hat{B}_{21} \\ \hat{B}_{12} \end{pmatrix} = \begin{pmatrix} \sum z_{ij} \hat{\alpha}_i \hat{\beta}_j \\ \sum z_{ij} \hat{\alpha}_i^2 \hat{\beta}_j \\ \sum z_{ij} \hat{\alpha}_i \hat{\beta}_j^2 \end{pmatrix} . \qquad (6.26)$$

An alternative series of tests can be obtained by extracting contrasts among the $\{z_{ij}\}$ on the basis of the ranks of the $\{\hat{\alpha}_i\}$ and $\{\hat{\beta}_j\}$. If for example $\{L(\alpha)_i\}$ represent the ranks of $\{\hat{\alpha}_i\}$ and $\{L(\beta)_j\}$ the ranks of $\{\beta_j\}$, then a valid test for non-additivity can be based on the statistic

$$\frac{(rc-r-c)\ T^*}{S-T^*}$$

where $S = \sum z_{ij}^2$ and $T^* = \dfrac{144\{\Sigma\ z_{ij}\ L(\alpha)_i\ L(\beta)_j\}^2}{r(r+1)(r-1)c(c+1)(c-1)}$.

D. A Numerical Example

The technique for constructing a test for non-additivity with more than one degree of freedom will be illustrated with the following set of data.

2.134	1.135	2.296	9.231	7.699	6.372
-2.172	.531	5.448	9.461	9.343	7.092
7.729	2.059	.371	6.354	8.257	5.944
4.462	5.610	7.507	1.090	9.658	5.352
3.960	6.897	7.582	9.705	7.109	11.719
4.319	8.121	9.146	10.352	7.613	9.151

This data was constructed, using the model

$$y_{ij} = \mu_{ij} + \frac{1}{150}\mu_{ij}^3 + e_{ij} \tag{6.27}$$

where $\mu_{ij} = \mu + \alpha_i + \beta_j$ and the $\{e_{ij}\}$ are NID(0,9) . The values for the

$\{\alpha_i\}$ are -2.0, -1.0, 0.0, .5, 1.0 and 1.5, for the $\{\beta_j\}$ are -2.0, -1.0, -.5, .5, 1.0 and 2.0 and $\mu = 5$. The mean of the 36 synthetic observations is 6.073. The excess over the true mean, which happens to be known in this case is a reflection of the non-additivity present. The estimated row effects are -1.262, -1.123, -.954, -.460, 1.756 and 2.044. The estimated column effects are -2.668, -2.014, -.681, 1.626, 2.207 and 1.532. The analysis of variance for the data is as follows:

Source of Variation	S.S.	d.f.	M.S.
Rows	67.407	5	13.481
Columns	128.989	5	25.798
Error	192.511	25	7.700
Total	388.907		

This analysis indicates a significant column effect. The residuals remaining after the removal of the mean, row effects and column effects are given in the following array.

- .009	-1.662	-1.835	2.794	.681	.029
-4.454	-2.405	1.179	2.885	2.186	.610
5.278	-1.046	-4.067	- .391	.931	- .707
1.517	2.011	2.575	-6.149	1.838	-1.793
-1.201	1.082	.434	.250	-2.927	2.358
-1.130	2.018	1.710	.609	-2.711	- .498

The sum of squares due to non-additivity for Tukey's test is equal to 8.549/7.665 = 1.115. However, the model

$$y_{ij} = \mu + \alpha_i + \beta_j + A_{11}\,\alpha_i\,\beta_j + A_{12}\,\alpha_i(\beta_j^2 - \beta^2) + A_{21}(\alpha_i^2 - \alpha^2)\beta_j + e_{ij}$$

leads to a test for non-additivity with 3 and 24 degrees of freedom. For this set of data the matrix equation given in (6.20) becomes

$$\begin{pmatrix} 241.50 & -99.32 & 205.39 \\ -99.32 & 297.68 & -84.47 \\ 502.39 & -84.47 & 234.53 \end{pmatrix} \begin{pmatrix} \hat{A}_{11} \\ \hat{A}_{12} \\ \hat{A}_{21} \end{pmatrix} = \begin{pmatrix} -45.44 \\ -61.03 \\ -\ .24 \end{pmatrix} .$$

The solutions for $\hat{A}_{11}$, $\hat{A}_{12}$ and $\hat{A}_{21}$ are -.861,508, -.310,400 and .641,692 respectively. The sum of squares due to non-additivity obtained by substituting into (6.19) is 57.937. The resulting F ratio with 3 and 22 degrees of freedom is 19.312/6.117 = 3.157. This value is significant at the 5% level, indicating the presence of non-additivity which was not detected by Tukey's test.

E. Power of Tests for Non-additivity

The power of Tukey's test for non-additivity has been studied rather extensively by Hogben (1963) and Ghosh and Sharma (1963). Also, it is a simple matter to verify that the statistics proposed in section C of this chapter are distributed as F ratios with appropriate degrees of freedom under the null hypothesis of additivity and normal independent errors. In general, the distributions under the alternative hypothesis of non-additivity appear to be intractable. An exception is the series of tests based on the ranks of the estimated marginal effects.

The procedure for computing the power of a test for non-additivity based on the ranks of the marginal totals will be developed by means of a specific example. However, the relatively simple extensions to more general cases will be pointed out at various stages.

Let $\{R(\alpha)_i\}$ be the r integers representing the ranks of the observed $\{\hat{\alpha}_i\}$. Similarly, let $\{R(\beta)_j\}$ be the c integers representing the ranks of the $\{\hat{\beta}_j\}$. A series of tests for non-additivity can be obtained by partitioning the error sum of squares into portions reflecting the regression on functions of the $\{R(\alpha)_i\}$ and $\{R(\beta)_j\}$ and a remainder. In particular, several sums of squares due to non-additivity which are independent under the hypothesis of additivity can be obtained from the regression of the residuals on the orthogonal polynomials based on the ranks. For example, let $\{L(r)_i\}$ and $\{Q(r)_i\}$ represent the linear and quadratic contrasts based on the ranks of the $\{\hat{\alpha}_i\}$ and $\{L(c)_j\}$ and $\{Q(c)_j\}$ the corresponding contrasts based on the ranks of the $\{\hat{\beta}_j\}$. A sum of squares with one degree of freedom, which will be inflated if non-additivity is present is

$$\frac{[\Sigma\, z_{ij}\, L(r)_i\, L(c)_j]^2}{[\Sigma\, L^2(r)_i][\Sigma\, L^2(c)_j]} \,. \tag{6.28}$$

Three other terms, each with one degree of freedom, are

$$\frac{[\Sigma\, z_{ij}\, L(r)_i\, Q(c)_j]^2}{[\Sigma\, L^2(r)_i][\Sigma\, Q^2(c)_j]} \,, \tag{6.29}$$

$$\frac{[\Sigma\, z_{ij}\, Q(r)_i\, L(c)_j]^2}{[\Sigma\, Q^2(r)_i][\Sigma\, L^2(c)_j]} \tag{6.30}$$

and

$$\frac{[\Sigma\ z_{ij}\ Q(r)_i\ Q(c)_j]^2}{[\Sigma\ Q^2(r)_i][\Sigma\ Q^2(c)_j]} \quad . \tag{6.31}$$

It is clear that the total error sum of squares can be broken down into a series of (r-1)(c-1) components which are all independent if an additive model holds.

The statement that (6.28) represents a sum of squares with one degree of freedom can be readily verified. Under the assumption of additivity the linear function

$$L = \sum_{ij} z_{ij}\ L(r)_i\ L(c)_j \tag{6.32}$$

has a normal distribution with mean zero and variance equal to

$$\sigma^2\ (\sum_{i=1}^{r} L^2(r)_i)(\sum_{j=1}^{c} L^2(c)_j) \quad .$$

The zero mean follows from the fact that the $\{z_{ij}\}$ have zero mean and are independent of the $\{L(r)_i\}$ and $\{L(c)_j\}$ (which are functions of the row and column effects). Normality follows from the normality of the residuals. The statement about the variance can be verified as follows:

$$\mathrm{Var}(\sum_{ij} z_{ij}\ L(r)_i\ L(c)_j\,|\{L(r)_i\}\ ,\ \{L(c)_j\}) \tag{6.33}$$

$$= \sum_{ij} L^2(r)_i\ L^2(c)_j\ \mathrm{Var}(z_{ij})\,|\{L(r)_i\}\ ,\ \{L(c)_j\}$$

$$+ \sum_{i\neq i'} \sum_{j} L(r)_i\ L(r)_{i'}\ L^2(c)_j\ \mathrm{Cov}(z_{ij}\ z_{i'j})\,|\ \{L(r)_i\}\ ,\ \{L(c)_j\}$$

$$+ \sum_i \sum_{j \neq j'} L^2(r)_i \, L(c)_j \, L(c)_{j'} \, \mathrm{Cov}(z_{ij} \; z_{ij'}) \,|\, \{L(r)_i\} \; , \; \{L(c)_j\}$$

$$+ \sum_{i \neq i'} \sum_{j \neq j'} L(r)_i \, L(r)_{i'} \, L(c)_j \, L(c)_{j'} \, \mathrm{Cov}(z_{ij} \; z_{i'j'}) \,|\, \{L(r)_i\} , \{L(c)_j\} .$$

Since the z_{ij} are independent of the marginal totals and hence the $\{L(r)_i\}$ and $\{L(c)_j\}$ it follows that the conditional variances are equal to the unconditional variances. Now.

$$\mathrm{Var}(z_{ij}) = \frac{(r-1)(c-1)}{rc} \sigma^2$$

$$\mathrm{Cov}(z_{ij}, z_{i'j}) = \frac{-(c-1)}{rc} \sigma^2 \qquad i \neq i'$$

$$\mathrm{Cov}(z_{ij}, z_{ij'}) = \frac{-(r-1)}{rc} \sigma^2 \qquad j \neq j'$$

and

$$\mathrm{Cov}(z_{ij}, z_{i'j'}) = \frac{1}{rc} \sigma^2 \qquad i \neq i' \; , \; j \neq j' \; .$$

Inserting these values into (6.33) leads to the desired result.

Similar justification can be advanced for (6.29), (6.30) and (6.31). It should be pointed out that among the rc residuals, a set of (r-1)(c-1) orthogonal linear contrasts with unit variance can be constructed. Under the hypothesis of an additive model and normal independent errors these have identical N(0,1) distributions. These can be examined by means of any one of several techniques to provide evidence of the tenability of this hypothesis. When non-additivity is present the expected value of L defined in (6.32) will not be zero. In fact, it will be

$$g\left(\sum_{i=1}^{r} \alpha_{(i)} L(r)_i\right)\left(\sum_{j=1}^{c} \beta_{(j)} L(c)_j\right)$$

where g is some constant and $\{\alpha_{(i)}\}$ and $\{\beta_{(j)}\}$ denote the $\{\alpha_i\}$ and $\{\beta_j\}$ respectively, in some order.

It follows that for fixed values of $\{L(r)_i\}$ and $\{L(c)_j\}$

$$\frac{N'}{\sigma^2} = \frac{\left[\sum_{ij} L(r)_i\, L(c)_j\, z_{ij}\right]^2}{\sigma^2\left[\sum_i L^2(r)_i\right]\left[\sum_j L^2(c)_j\right]} \tag{6.34}$$

has a non-central χ^2-distribution with one degree of freedom and non-centrality parameter equal to

$$\frac{g^2}{\sigma^2}\lambda_1 = \frac{g^2}{\sigma^2}\frac{\left[\sum_i L(r)_i\, \alpha_i\right]^2\left[\sum_j L(c)_j\, \beta_j\right]^2}{\left[\sum_i L^2(r)_i\right]\left[\sum_j L^2(c)_j\right]} \quad . \tag{6.35}$$

Let

$$E = \sum_{ij} z_{ij}^2$$

be the error sum of squares under the additive model. It is shown by Ghosh and Sharma (1963) that E/σ^2 has a non-central χ^2-distribution with (r-1)(c-1) degrees of freedom and non-centrality parameter $g^2 \sum_i \alpha_i^2 \sum_j \beta_j^2/\sigma^2$, under the model $y_{ij} = \mu + \alpha_i + \beta_j + g\, \alpha_i\, \beta_j + e_{ij}$ and normal, identically distributed errors. This holds regardless of whether the marginal values are fixed or not. It follows that $(E-N')/\sigma^2$ has a non-central χ^2-distribution with (rc-r-c) degrees of freedom and non-centrality parameter

$$\frac{g^2}{\sigma^2}\lambda_2 = \frac{g^2}{\sigma^2}\left\{ \sum_i \alpha_i^2 \sum_j \beta_j^2 - \frac{\left[\sum_i L(r)_i\, \alpha_i\right]^2 \left[\sum_j L(c)_j\, \beta_j\right]^2}{\left[\sum_i L^2(r)_i\right]\left[\sum_j L^2(c)_j\right]} \right\} \tag{6.36}$$

for fixed $\{L(r)_i\}$ and $\{L(c)_j\}$. It follows immediately from Schwartz's inequality that this expression will always be positive.

Now form the statistic

$$T = \frac{N'(rc-r-c)}{E-N'} . \tag{6.37}$$

Since the numerator and denominator are independent for fixed $\{L(r)_i\}$ and $\{L(c)_j\}$ it follows that conditionally T has a doubly non-central F-distribution with one and (rc-r-c) degrees of freedom and non-centrality parameters $\frac{g^2\,\lambda_1}{\sigma^2}$ and $\frac{g^2\,\lambda_2}{\sigma^2}$ when $g\neq 0$. When $g=0$, T has a central F-distribution with one and (rc-r-c) degrees of freedom. The hypothesis that $g=0$ is rejected if the realized value of the statistic is larger than some prechosen upper tail percentage point of the F-distribution with one and (rc-r-c) degrees of freedom.

The statistic

$$t = T/(rc-r-c)$$

has the probability density

$$\exp\left\{-\frac{g^2}{2\sigma^2}\sum_i \alpha_i^2 \sum_j \beta_j^2\right\} \sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\left(\frac{g^2}{2\sigma^2}\right)^{p+q} \cdot \tag{6.38}$$

$$\frac{\lambda_1^p \lambda_2^q}{p!\ q!\ B\{\tfrac{1}{2}+p,\ \tfrac{1}{2}(rc-r-c)+q\}} \cdot \frac{t^{p-\frac{1}{2}}}{(1+t)^{\frac{1}{2}(r-1)(c-1)+p+q}}$$

for $0 \leq t \leq \infty$ and fixed $\{L(r)_i\}$ and $\{L(c)_j\}$. $B\{n,m\}$ is the beta function. The unconditional distribution of t can now be obtained by multiplying the above density by the density of $\{L(r)_i\}$ and $\{L(c)_j\}$ and then integrating to obtain the marginal density of t .

From the manner in which $\{L(r)_i\}$ and $\{L(c)_j\}$ were defined it follows that these sets can consist of only r! and c! sets of values respectively corresponding to the possible rankings of the r and c marginal means. Hence, the marginal density function of t is obtained by summing over the probabilities with which the various sets $\{L(r)_i\}$ and $\{L(c)_j\}$ can occur.

It can be seen from (6.38) that the density of t depends on $\{L(r)_i\}$ and $\{L(c)_j\}$ only through the values of λ_1 and λ_2 . Hence if the conditional density of t is denoted by

$$F^*(g,\ \sigma^2,\ t|\lambda_{1k}\ ,\ \lambda_{2h})$$

then the unconditional density can be denoted by

$$F(g,\ \sigma^2,\ t) = \sum_{k=1}^{r!}\sum_{h=1}^{c!} F^*(g,\ \sigma^2,\ t|\lambda_{1k},\lambda_{2h})\ f(\lambda_{1k}\ ,\ \lambda_{2h}) \tag{6.39}$$

where $f(\lambda_{1k}, \lambda_{2h})$ denotes the probability with which the k-th and the h-th realizations $\{L(r_i)\}$ and $\{L(c)_j\}$ occur i.e., the probability with which the given arrangement of marginal means will occur. The power of the test for a given value of g is then given by the integral

$$\int_{t_0}^{\infty} F(g, \sigma^2, t)\, dt$$

where t_0 is chosen so that the desired level of significance is achieved by the test. Since the summations in (6.39) involve only a finite number of terms, these summations can be carried out prior to the integration. From the nature of

$$F^*(g, \sigma^2, t \mid \lambda_{1k}, \lambda_{2h})$$

we see that this amounts to evaluating the moments

$$E\left[\lambda_1^m \lambda_2^n\right] = \left[E\left(\frac{g^2}{\sigma^2}\right)^{m+n} \left\{\frac{\left[\sum_i L(r)_i\, \alpha_i\right]^2}{\left[\sum_i L^2(r)_i\right]} \times \frac{\left[\sum_j L(c)_j\, \beta_j\right]^2}{\left[\sum_j L^2(c)_j\right]}\right\}^m \right.$$

$$\left. \left\{\sum_i \alpha_i^2 \sum_j \beta_j^2 - \frac{\left[\sum_i L(r)_i\, \alpha_i\right]^2}{\left[\sum_i L^2(r)_i\right]} \times \frac{\left[\sum_j L(c)_j\, \beta_j\right]^2}{\left[\sum_j L^2(c)_j\right]}\right\}^n \right]$$

for various values of m and n. It follows that

$$F(g, \sigma^2, t) = \exp\left\{-\frac{g^2}{2\sigma^2} \sum_i \alpha_i^2 \sum \beta_j^2\right\} \times \sum_{p,q=0}^{\infty} \left(\frac{g^2}{2\sigma^2}\right)^{p+q} \times \tag{6.41}$$

$$\frac{E\left[\lambda_1^p \lambda_2^q\right]}{p!q!} \times \frac{t^{p-\frac{1}{2}}}{(1+t)^{\frac{1}{2}(r-1)(c-1) + p + q}} \times \frac{1}{B\{\frac{1}{2} + p, \frac{1}{2}(r-1)(c-1) + q\}}$$

for $0 \leq t \leq \infty$. This is recognized as a series of beta functions with appropriate weights.

F. Numerical Examples of Power Calculations

Under the assumption of normally distributed errors the probability of obtaining a given $\{L(r)_i\}$ is simply the probability that the r row means will have the appropriate ranks. These means are independent with means α_i , i=1, 2, ..., r and variance σ^2/c . The probability of a given $\{L(c)_j\}$ is obtained in the same manner.

The actual numerical methods needed to obtain the values for the multivariate normal integrals are outlined in the appendix. Unfortunately the values obtained depend on the true row and column effects. Once these probabilities have been obtained it is a relatively simple matter to obtain the values for $E[\lambda_1^m \lambda_2^n]$.

If t_0 denotes the value which t must exceed in order to yield a test of appropriate size then the power of the test for given values of σ^2 and g is defined as

$$\int_{t_0}^{\infty} F(g, \sigma^2, t)\, dt = \exp\left\{-\frac{g^2}{2\sigma^2} \sum_i \alpha_i^2 \sum_j \beta_j^2\right\} \times \sum_{p,q=0}^{\infty} \left(\frac{g^2}{2\sigma^2}\right)^{p+q} \times \frac{E[\lambda_1^p \lambda_2^q]}{p!q!} \times$$

$$\int_{t_0}^{\infty} \frac{t^{p-\frac{1}{2}}}{(1+t)^{\frac{1}{2}(r-1)(c-1) + p + q}} \times \frac{1}{B\{\frac{1}{2} + p, \frac{1}{2}(rc-r-c) + q\}}\, dt \, . \tag{6.42}$$

The numerical values for these integrals can be obtained from Pearson's tables of the incomplete beta function if one makes the transformation $u = \frac{t}{1+t}$ and obtains

$$\int_{t_0}^{\infty} \frac{t^{p-\frac{1}{2}}}{(1+t)^{\frac{1}{2}(r-1)(c-1) + p + q}}\, dt = \int_{u_0}^{1} u^{p-\frac{1}{2}} (1-u)^{\frac{1}{2}(rc-r-c)+q-1}\, du \, . \tag{6.43}$$

Since a large number of incomplete beta values were needed and rather extensive interpolation in the table was required it was decided to use a quick method of computing the necessary values given by Müller (1930). The accuracy of the method was checked by evaluating the incomplete beta function at a grid of points in the table in the region of interest and then comparing the tabular and computed values. The method proved to be quite satisfactory. The details of the method are given in the appendix.

A section of the power curve for a test of non-additivity for a 5 × 5 table was evaluated and is given in Figure 1. The main effects used to compute the curve were

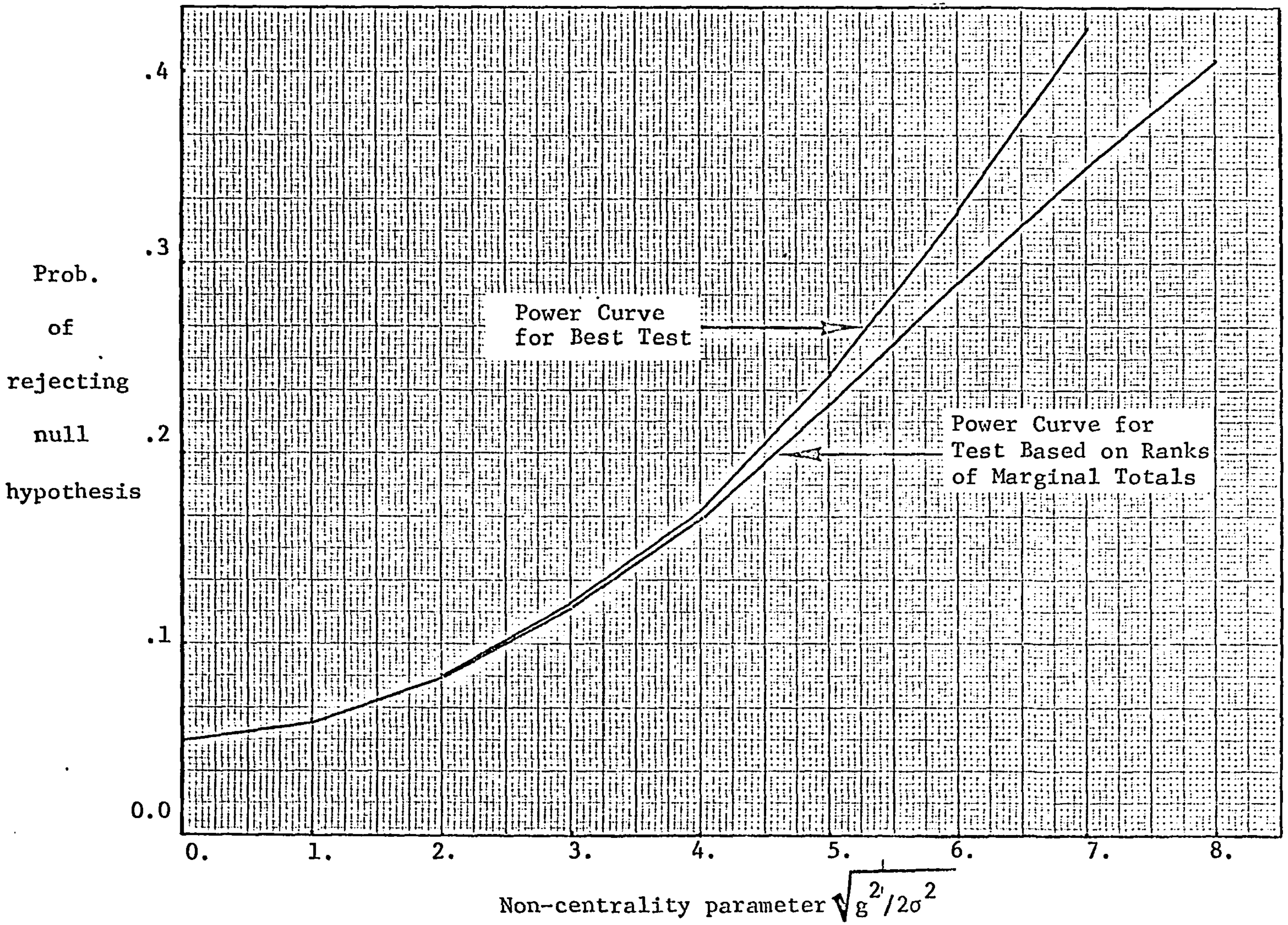

Figure 1. Power Curve for Test of Non-Additivity Based on Ranks in a 5 × 5 Table

$$\alpha_1 = -\ 1.16 = \beta_1$$

$$\alpha_2 = -\ \ .5 = \beta_2$$

$$\alpha_3 = \ \ 0 = \beta_3$$

$$\alpha_4 = \ \ .5 = \beta_4$$

$$\alpha_5 = \ \ 1.16 = \beta_5 \ .$$

A value of $\sigma^2 = 2.5$ and a 5% significance level were used. Only a certain portion of the power curve is given since the nature of the methods of computation is such that many more terms must be retained in some of the approximations in order to yield accurate results for $\frac{g^2}{2\sigma^2}$ larger than .6. Since the expense of computing would increase considerably if more terms were retained and since the portion of the curve displayed was thought to be the most important, it was decided to terminate the calculations.

For the sake of comparison the power curve for "best" test for non-additivity was also calculated. This same "best" test was also used by Hogben (1963) to provide a standard by which to evaluate Tukey's test for non-additivity. This "best" test is motivated by the following considerations:

In the framework of the model

$$y_{ij} = \mu + \alpha_i + \beta_j + g\ \alpha_i\ \beta_j + e_{ij}$$

and the e_{ij} are independent $N(0,\sigma^2)$ Tukey's test for non-additivity is a test of the composite hypothesis that $g = 0$. A natural idealization of this is the case when all the parameters in the model except g are known. In this case the model would be rewritten as

$$z_{ij} = y_{ij} - \mu - \alpha_i - \beta_j \tag{6.44}$$

$$= g\,\alpha_i\,\beta_j + e_{ij} .$$

The least squares estimate of g is

$$\hat{g} = \frac{\sum_{ij} z_{ij}\,\alpha_i\,\beta_j}{\Sigma\,\alpha_i^2\,\Sigma\,\beta_j^2} . \tag{6.45}$$

In this case the z_{ij} are independent normal with mean $g\,\alpha_i\,\beta_j$ and variance σ^2. Therefore $\hat{g}/\sigma^2$ is normally distributed with mean g/σ^2 and variance $\frac{1}{\sum_i \alpha_i^2 \sum_j \beta_j^2}$. It follows that $\frac{\hat{g}}{\sigma^2}\sqrt{\Sigma\,\alpha_i^2\,\Sigma\,\beta_j^2}$ has mean $\frac{g}{\sigma}\sqrt{\Sigma\,\alpha_i^2\,\Sigma\,\beta_j^2}$ and unit variance. The power of the best test obtained in this manner is equal to

$$\Pr\left\{\left|\frac{\hat{g}}{\sigma}\sqrt{\Sigma\,\alpha_i^2\,\Sigma\,\beta_j^2}\right| > T_0 \,\middle|\, g\right\} \tag{6.46}$$

where T_0 is chosen so that

$$\Pr\left\{\left|\frac{\hat{g}}{\sigma}\sqrt{\Sigma\,\alpha_i^2\;\Sigma\,\beta_j^2}\right| > T_0 \,|\, g = 0\right\}$$

is equal to some predetermined value. This probability can be evaluated easily using tables of the cumulative normal. It may be noted that as r and c increase the power of this "best" test will improve since $\sqrt{\Sigma\,\alpha_i^2}$ and $\sqrt{\Sigma\,\beta_j^2}$ will both tend to increase. An examination of Figure 1 indicates that the performance of the test based on the ranks of the marginal totals is very satisfactory.

A more general test for non-additivity based on the ranks of the means will now be discussed. This test is motivated by the same considerations as the test for non-additivity based on a sum of squares with more than one degree of freedom for non-additivity. The discussion will again be in terms of an example with indications of the modifications needed for further extensions. Consider first replacing $\{L(r)_i\}$ by $\{Q(r)_i\}$ where $\{Q(r)_i\}$ is the set of coefficients for orthogonal polynomials of second degree for r observed values. The second degree polynomial coefficients were chosen in this case for the sake of definiteness. Third, fourth or even higher order polynomial coefficients could be chosen provided they are defined and appear realistic.

Now consider the linear function of the residuals

$$\sum_{ij} L(c)_j\; Q(r)_i\; z_{ij}\;.$$

Since the $\{z_{ij}\}$ have a normal distribution it follows that the above contrast also has a normal distribution. The covariance between the two functions

$$\sum_{ij} L(r)_i \, L(c)_j \, z_{ij} \quad \text{and} \quad \sum_{ij} Q(r)_i \, L(c)_j \, z_{ij}$$

is equal to

$$\sum_{ii'} \sum_{jj'} L(r)_i \, Q(r)_{i'} \, L(c)_j \, L(c)_{j'} \, \mathrm{Cov}(z_{ij} \; z_{i'j'})$$

$$= \sum_j L(c)_j \, L(c)_j \, \{ \sum_{ii'} L(r)_i \, Q(r)_{i'} \, \mathrm{Cov}(z_{ij} \; z_{i'j}) \}$$

$$+ \sum_{j \neq j'} L(c)_j \, L(c)_{j'} \, \{ \sum_{ii'} L(r)_i \, Q(r)_{i'} \, \mathrm{Cov}(z_{ij} \; z_{i'j'}) \}$$

$$= \sum_j L(c)_j \, L(c)_j \, \{ \sum_i L(r)_i \, Q(r)_i \, \frac{(r-1)(c-1)}{rc} \, \sigma^2 \tag{6.47}$$

$$- \sum_{i \neq i'} L(r)_i \, Q(r)_{i'} \frac{(c-1)}{rc} \, \sigma^2 \}$$

$$+ \sum_j L(c)_j \, L(c)_j \, \{ \sum_i L(r)_i \, Q(r)_{i'} \frac{(r-1)}{rc} \, \sigma^2$$

$$- \sum_{i \neq i'} L(r)_i \, Q(r)_i \, \frac{1}{rc} \, \sigma^2 \}$$

$$= 0$$

since $\sum_i L(r)_i = \sum_i Q(r)_i = \sum_i L(r)_i \, Q(r)_i = 0$.

Hence the two functions are independent. By the same type of argument it can be shown that any pair of functions

$$\sum_{ij} A(r)_i \; B(c)_j \; z_{ij} \quad \text{and} \quad \sum_{ij} A'(r)_i \; B'(c)_j \; z_{ij}$$

are independent if either $\{A(r)_i\}$ and $\{A'(r)_i\}$ are the coefficients for the orthogonal polynomials for r points are of unequal degree or $\{B(c)_j\}$ and $\{B'(c)_j\}$ are the coefficients for the orthogonal polynomials for c points are of unequal degree. It follows that under the same conditions the two sums of squares

$$\frac{\left[\sum_{ij} A(r)_i \; B(c)_j \; z_{ij}\right]^2}{\sigma^2 \left[\sum_i A^2(r)_i\right]\left[\sum_j B^2(c)_j\right]}$$

and

$$\frac{\left[\sum_{ij} A'(r)_i \; B(c)_j \; z_{ij}\right]^2}{\sigma^2 \left[\sum_i A'^2(r)_i\right]\left[\sum_j B^2(c)_j\right]}$$

are independently distributed with non-central χ^2-distributions with one degree of freedom each and non-centrality parameters

$$\frac{g^2}{\sigma^2} \; \frac{\left[\sum_i A(r)_i \; \alpha_i\right]^2}{\left[\sum_i A^2(r)_i\right]} \times \frac{\left[\sum_j B(c)_j \; \beta_j\right]^2}{\left[\sum_j B^2(c)_j\right]}$$

and

$$\frac{g^2}{\sigma^2}\ \frac{\left[\sum\limits_i A'(r)_i\ \alpha_i\right]^2}{\left[\sum\limits_i A'^2(r)_i\right]} \times \frac{\left[\sum\limits_j B(c)_j\ \beta_j\right]^2}{\left[\sum\limits_j B^2(c)_j\right]}$$

respectively for fixed ranking of the marginal means. These non-centrality parameters are computed under the assumption of a model of the type $z_{ij} = \mu + \alpha_i\ \ \beta_j + g\ \alpha_i\ \beta_j + e_{ij}$. By this device the error sum of squares can be broken up into a total of (r-1)(c-1) independent sums of squares.

Since these sums of squares are independent they can be pooled at will and the non-centrality parameters of the corresponding χ^2-distribution will add accordingly. Consequently one can obtain a sum of squares due to non-additivity with any number of degrees of freedom $\leq$ (r-1)(c-1) .

A section of the power curve for a test of non-additivity of the type

$$y_{ij} = \mu + \alpha_i + \beta_j + g\ \alpha_i\ \beta_j + e_{ij}$$

$$i = 1, 2, \ldots, 5,\ j = 1, 2, \ldots, 5$$

with 4 and 12 degrees of freedom has been computed. The main effects for this example were $\alpha_1 = \beta_1 = -1.16$, $\alpha_2 = \beta_2 = -.5$, $\alpha_3 = \beta_3 = 0$, $\alpha_4 = \beta_4 = .5$, $\alpha_5 = \beta_5 = 1.16$. σ^2 was equal to 2.5 and a 5% level of test was used. The four sums of squares which were pooled to form the sum of squares due to non-additivity were

$$\frac{\left[\sum_{ij} z_{ij}\ L(r)_i\ L(c)_j\right]^2}{\left[\sum_i L^2(r)_i\ \sum_j L^2(c)_j\right]}$$

$$\frac{\left[\sum_{ij} z_{ij}\ L(r)_i\ Q(c)_j\right]^2}{\left[\sum_i L^2(r)_i\right]\left[\sum_j Q(c)_j\right]}$$

$$\frac{\left[\sum_{ij} z_{ij}\ Q(c)_i\ L(r)_j\right]^2}{\left[\sum_i Q^2(r)_i\right]\left[\sum_j L^2(r)_j\right]}$$

and

$$\frac{\left[\sum_{ij} z_{ij}\ Q(r)_i\ Q(c)_j\right]^2}{\left[\sum_i Q^2(r)_i\right]\left[\sum_j Q^2(c)_j\right]}$$

$\{L(r)_i\}$ and $\{L(c)_j\}$ are the coefficients for the first degree orthogonal polynomial values for r and c points respectively. $\{Q(r)_i\}$ and $\{Q(c)_j\}$ are the coefficients for the second degree orthogonal polynomial values. The non-centrality parameter for $\frac{\Lambda}{\sigma^2}$ = (sum of squares due to non-additivity)/σ^2 was

$$\frac{g^2}{\sigma^2}\,\Lambda = \frac{g^2}{\sigma^2}\left[\frac{\left[\Sigma\, L(r)_i\, \alpha_i\right]^2}{\left[\Sigma\, L^2(r)_i\right]}\cdot\frac{\left[\Sigma\, L(c)_j\, \beta_j\right]^2}{\left[\Sigma\, L^2(c)_j\right]} + \frac{\left[\Sigma\, L(r)_i\, \alpha_i\right]^2}{\left[\Sigma\, L^2(r)_i\right]}\,\frac{\left[\Sigma\, Q(c)_j\, \beta_j\right]^2}{\left[\Sigma\, Q^2(c)_j\right]}\right.$$

$$\left. + \frac{\left[\Sigma\, Q(r)_i\, \alpha_i\right]^2}{\left[\Sigma\, Q^2(r)_i\right]}\,\frac{\left[\Sigma\, L(c)_j\, \beta_j\right]^2}{\left[\Sigma\, L^2(c)_j\right]} + \frac{\left[\Sigma\, Q(r)_i\, \alpha_i\right]^2}{\left[\Sigma\, Q^2(r)_i\right]}\,\frac{\left[\Sigma\, Q(c)_j\, \beta_j\right]^2}{\left[\Sigma\, Q^2(c)_j\right]}\right].$$

The test statistic in this case is

$$T = \frac{N^*(r-1)(c-1) - 4}{4(E-N^*)}$$

where $E = \sum_{ij} (y_{ij} - y_{i.} - y_{.j} + y_{..})^2$.

The statistic $t = \dfrac{N^*}{E-N^*}$ has probability density

$$\exp\left\{-\frac{g^2}{2\sigma^2}\sum \alpha_i^2 \sum \beta_j^2\right\} \sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\left(\frac{g^2}{2\sigma^2}\right)^{p+q} \frac{E\left[\Lambda^p(\Sigma\, \alpha_i^2\, \Sigma\, \beta_j^2 - \Lambda)^q\right]}{p!q!\; B\{2+p,\ \tfrac{1}{2}(r-1)(c-1)-2+q\}}\cdot$$

$$\frac{t^{p+1}}{(1+t)^{\frac{1}{2}(r-1)(c-1) + p + q}}$$

A graph of the power curve is given in Figure 2 . The power curve for the ideal test given in Figure 1 is repeated here even though the comparison is not strictly fair. Also the model used for the computations is appropriate only for a test for non-additivity with one degree of freedom, based only on the first degree orthogonal polynomials, since the non-additivity is contributed by the term $\bar{g}\, \alpha_i\, \beta_j$. If the non-additivity had been contributed by a term like $g\,\alpha_i\,\beta_j + h\,\alpha_i^2\,\beta_j + k\,\alpha_i\,\beta_j^2 + m\,\alpha_i^2\,\beta_j^2$, then the four degree of freedom test would have been appropriate, and would have shown up more favorably. In a sense the extra degrees of freedom weaken the test by diluting the effect of increasing g .

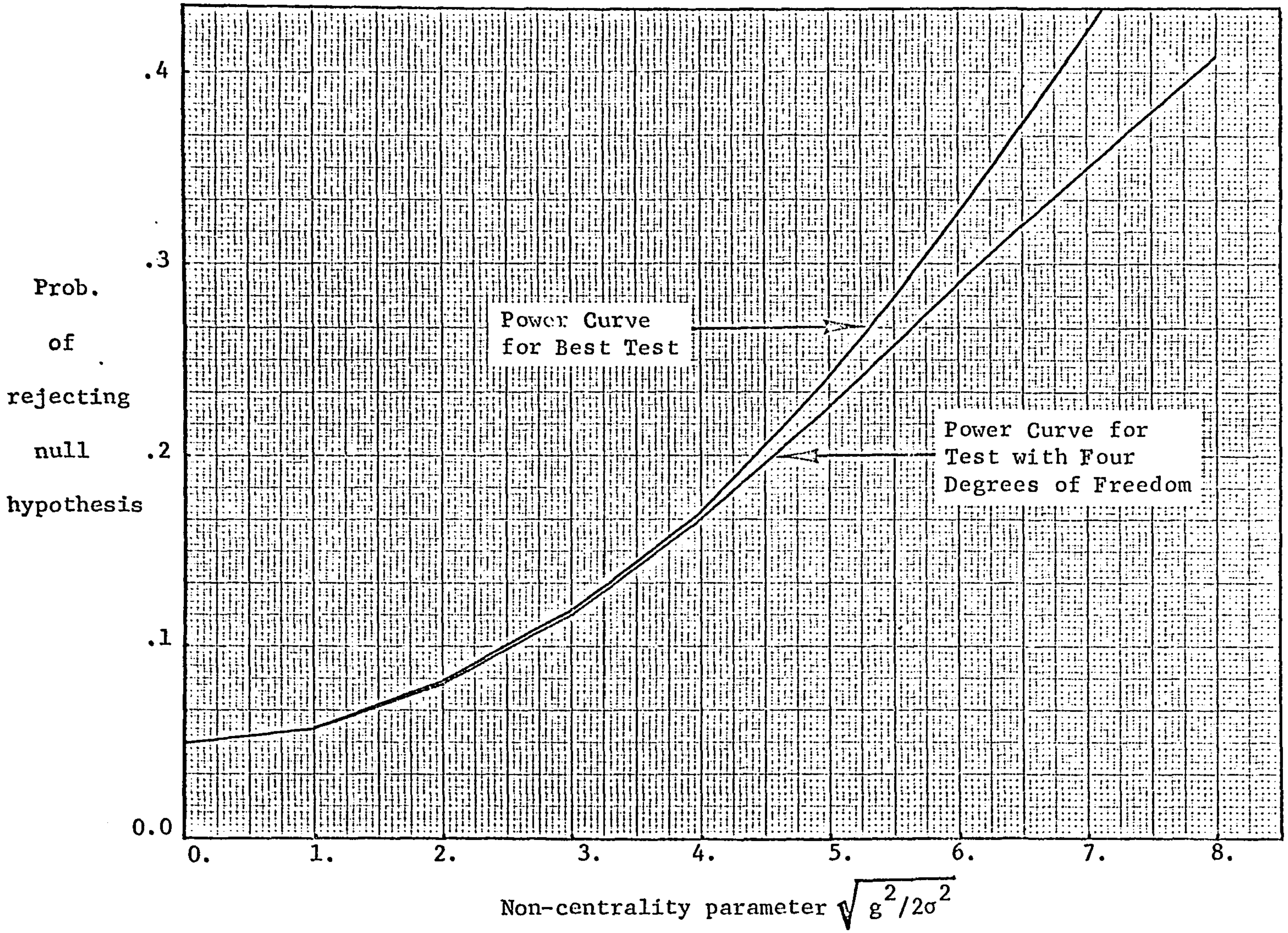

Figure 2. Power Curve for Test of Non-Additivity with Four Degrees of Freedom

VII. ESTIMATING TRANSFORMATIONS

A. Introduction

It is common practice in the analysis of data to assume that the observations are independent and normally distributed with constant variance and expected value specified by a model which is linear in the parameters. In the previous chapters techniques for examining the residuals, with the aim of detecting deviations from these assumptions were developed. The emphasis in this chapter is on finding transformations which will modify the observed data in an attempt to satisfy the assumptions. The general notion is to restrict attention to transformations indexed by unknown parameters and then estimate these by the method of maximum likelihood. The emphasis is on the family of transformations given by

$$y_i = \ln(x_i + \alpha) \tag{7.1}$$

though some mention is made of the more general family

$$y_i = (x_i + \alpha)^{\beta} \; . \tag{7.2}$$

Of course the first can be shown to be a special limiting case of the latter, and the two treated as one family. However, this will not be done in this thesis.

B. Review of Previous Work

One of the earliest writers to attempt to solve the problem posed in this chapter was Cohen (1951). The approach taken by this author was to write the likelihood as

$$L(\alpha,\mu,\sigma) = \begin{cases} (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{n} (x_i + \alpha)^{-1} \exp\left\{- \tfrac{1}{2}\left(\frac{\ln(x_i+\alpha)-\mu}{\sigma}\right)^2\right\} \\ \text{if all } x_i > -\alpha, \quad \sigma > 0 \text{ and } -\infty < \mu < \infty \\ 0 \text{ otherwise }, \end{cases} \tag{7.3}$$

compute the partial derivatives with respect to the three unknown parameters and then use an iterative technique to solve the resulting non-linear equations. Hill (1963) showed that one need not arrive at useful estimates of the parameters by this method, since there exists a path in the three dimensional parameter space, along which $L(\alpha,\mu,\sigma)$ continues to increase, regardless of the observed $\{x_i\}$. Hill concluded by showing that more suitable estimates of the parameters could be obtained if one assumed appropriate prior distributions for the parameters and invoked a Bayesian type argument. The prior distribution is chosen so that parameter values in the area where the likelihood becomes unbounded are eliminated from consideration.

The major empahsis in Hill's paper however, is to estimate the parameters in the three parameter lognormal distribution rather than to find a transformation. In his formulation the α in (7.3) represented a threshold for the phenomenon being studied.

In a recent paper, in which the emphasis was on finding a transformation, Box and Cox (1964) failed to recognize the problem of an unbounded likelihood. An approach suggested by these authors was to carry out the maximization of the likelihood in two steps. The first step was based on the fact that for a fixed α , (7.3) is maximized by equating μ and σ to the mean and standard deviation of the $\{y_i\}$ defined by the equations

$$y_i = \ln(x_i + \alpha) \text{ for } i = 1, 2, \ldots n .$$

These values, however, depend on α and must be denoted by $\hat{\mu}(\alpha)$ and $\hat{\sigma}(\alpha)$. Substituting these into (7.3), and taking logarithms leads to

$$\ln\left(L(\alpha,\hat{\mu}(\alpha), \hat{\sigma}(\alpha))\right)$$

$$= \text{constant} - n \ln\left(\hat{\sigma}(\alpha)\right) - \sum_{i=1}^{n} \ln\left(x_i+\alpha\right) . \tag{7.4}$$

The second step consists of finding the value for α which maximizes (7.4). The suggested procedure is to evaluate and plot $\ln\left(L(\alpha,\hat{\mu}(\alpha),\hat{\sigma}(\alpha))\right)$ for series of values of α , and then read off the maximizing value for $\hat{\alpha}$. Unfortunately the authors failed to realize that, as shown by Hill (1963), (7.4) becomes infinitely large as $-\alpha$ approaches x_{min} . However, it often happens that there exists a value for α , such that $\ln\left(L(\alpha,\hat{\mu}(\alpha), \hat{\alpha}(\alpha))\right)$ is at a local maximum. This fact was exploited in a recent paper by Harter and Moore (1966). They proposed an iterative method to solve the non-linear equations obtained by computing the partial derivatives of the logarithm of (7.3) with respect to each of the three parameters. They state that a local maximum may fail to exist, especially when dealing with small samples. When this happens they suggest that the smallest observed value (and those equal to it) be discarded and the iterative procedure repeated. However, $\hat{\alpha}$ is still restricted to being less than all the observed $\{x_i\}$. Consequently the likelihood is bounded and a maximum attained.

A common feature of all of the references cited above is that the analysis is performed as though the observations are perfect, i.e., recorded without error. It has recently been argued by Kempthorne (1966) that all observations are subject to a grouping error, actually specified by the scientist and hence are in fact discrete. It follows that a properly defined likelihood function is a bounded function of whatever unknown parameters are involved. Also, the powerful theorems concerning maximum likelihood estimation in the case of observations from a multinomial distribution apply.

If Δ represents the size of the intervals into which the data are grouped in the course of the experiment and it is assumed that the underlying distribution is the three parameter lognormal then the correct form of the likelihood is

$$L(\alpha,\mu,\sigma) = \begin{cases} K \prod_i (p_i)^{f_i} \\ \text{if } -\alpha < (i'+\tfrac{1}{2})\Delta \text{ where } i' \text{ is the min. } i \\ \text{for which } f_i \neq 0 \\ 0 \text{ otherwise ,} \end{cases} \tag{7.5}$$

where K is a constant,

$$p_i = \begin{cases} \displaystyle\int_{\max((i-\frac{1}{2})\Delta,\ -\alpha)}^{(i+\frac{1}{2})\Delta} \frac{1}{\sqrt{2\pi}\sigma} \exp - \tfrac{1}{2}\left\{\frac{\ln(x+\alpha) - \mu}{\sigma}\right\}^2 \frac{dx}{(x+\alpha)} \\ \text{where } -\alpha < (i+\tfrac{1}{2})\Delta \ , \ \sigma > 0 \text{ and } -\infty < \mu < \infty \\ 0 \text{ otherwise,} \end{cases}$$

and f_i is the number of observations recorded as equal to Δi .

It is obvious from the above definition that for a fixed grouping error, the $\{p_i\}$ are functions of the parameters α, μ and σ . Since $p_i \leq 1$ for all i , it follows that $L(\alpha,\mu,\sigma)$ defined by (7.5) is a bounded, continuous function of the three parameters. Consequently, it can be maximized. It is shown by Kendall and Stuart (1961) that the values which maximize the likelihood function are consistent estimates of the unknown parameters. For small sample sizes, however, there is no guarantee that the solution will be unique, though it has been shown by Huzurbazar (1948) that ultimately, as n increases, there is a unique consistent solution. Also it has been shown by Cramér (1946) that if the first two partial derivatives of the likelihood function exist in an interval about the true parameter values and the expected value of matrix of second partials exists and is non-singular then, the maximum likelihood estimates are asymptotically multivariate normal with variance-covariance matrix equal to the negative of the inverse of the expected value of the matrix of second partials.

C. The Likelihood when Grouping Errors are Ignored

The likelihood, as one is led to define it, when grouping errors are ignored will be examined in some detail in this section. For a simple random sample of n observations, the likelihood under these conditions is written as

$$L(\alpha,\mu,\sigma) = \begin{cases} (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{n} (x_i+\alpha)^{-1} \exp - \left\{ \tfrac{1}{2}\left(\dfrac{\ln(x_i+\alpha)-\mu}{\sigma}\right)^2 \right\} \\ \quad \text{if all } x_i > -\alpha \text{ , } \sigma > 0, \ -\infty < \mu < \infty \\ 0 \text{ if otherwise.} \end{cases} \tag{7.6}$$

This function has been examined in considerable detail by Hill (1963). For example, he points out that this function approaches zero as any one of the parameters is allowed to approach the extreme values in its range, i.e., μ approaches $\pm\infty$, σ approaches 0 or $+\infty$ or α approaches $-x_{min}$ or $-\infty$, while the other two parameters are held constant. However, he also shows that there exists a path $(\alpha,\hat{\mu}(\alpha)\ ,\ \hat{\sigma}(\alpha))$ along which $L(\alpha,\hat{\mu}(\alpha)\ ,\ \hat{\sigma}(\alpha))$ as defined by (7.6) goes to infinity. The two quantities $\hat{\mu}(\alpha)$ and $\hat{\sigma}(\alpha)$ are the values for μ and σ which maximize (7.6) for a fixed α . These are simply the mean and standard deviation of the transformed observations. It is a simple matter to show that

$$\begin{aligned} &\ln\Big(L(\alpha,\hat{\mu}(\alpha)\ ,\ \sigma(\alpha))\Big) \\ &\quad = K - n[\hat{\mu}(\alpha) + \ln(\hat{\sigma}(\alpha))] \end{aligned} \tag{7.7}$$

where K is a constant and can be ignored.

In order to examine some features of the likelihood in more detail, under ideal conditions, an artificial sample was constructed with $\alpha = -3$, $\mu = 1.5$ and $\sigma = .5$. The 'observations' were assumed to be accurate to the nearest unit and were recorded as 3.5, 4.5, ..., 27.5 . These data are reproduced in Table 6. The frequency assigned to each

Table 6. Artificial Data Sample from the

Three-Parameter Lognormal with $\alpha = -3$, $\mu = 1.5$ and $\sigma = .5$

Value	Frequency	Value	Frequency
3.5	13	16.5	53
4.5	524	17.5	35
5.5	1582	18.5	24
6.5	1971	19.5	16
7.5	1781	20.5	11
8.5	1319	21.5	8
9.5	943	22.5	5
10.5	637	23.5	4
11.5	407	24.5	3
12.5	286	25.5	2
13.5	178	26.5	1
14.5	115	27.5	1
15.5	78	Total	9997

value is proportional to the probability of a true observation being in the interval of unit length about the recorded value.

For this sample, (7.7) has a relative maximum at $\hat{\alpha} = -2.7$, $\hat{\mu} = 1.57$ and $\hat{\sigma}^2 = .224$. These values clearly do not agree with true values which are known to be $\alpha = -3$, $\mu = 1.5$ and $\sigma^2 = .25$. It is interesting to note that while (7.7) is not maximized at $\alpha = -3$, the corresponding values for μ and σ^2 at that point are 1.496 and .2598.

The behavior of the likelihood in the neighborhood of the relative maximum was examined by computing the quantity

$$-9997[\hat{\mu}(\alpha) + \ln \hat{\sigma}(\alpha)]$$

for α ranging from -3.4 to -1.6 by steps of .1 . These computations are summarized in Table 7. Clearly the method advocated by Harter and Moore (1966) would select the estimates $\hat{\alpha} = -2.7$, $\hat{\mu} = 1.57$ and $\hat{\sigma} = .473$.

Table 7. Likelihood in the Neighborhood of the Relative Maximum when Grouping Errors are Ignored

α	$-9997[\hat{\mu}(\alpha)+\ln\hat{\sigma}(\alpha)]$	$\hat{\mu}(\alpha)$	$\hat{\sigma}^2(\alpha)$
-3.4	-8372.	1.3851	.33446
-3.3	-8293.	1.4147	.31027
-3.2	-8255.	1.4429	.29101
-3.1	-8232.	1.4699	.27444
-3.0	-8217.	1.4959	.25978
-2.9	-8209.	1.5211	.24661
-2.8	-8205.	1.5455	.23468
-2.7	-8203.	1.5691	.22378
-2.6	-8205.	1.5921	.21377
-2.5	-8207.	1.6144	.20455
-2.4	-8211.	1.6361	.19601
-2.3	-8216.	1.6573	.18807
-2.2	-8222.	1.6779	.18068
-2.1	-8229.	1.6981	.17378
-2.0	-8237.	1.7178	.16732
-1.9	-8244.	1.7370	.16126
-1.8	-8253.	1.7559	.15556
-1.7	-8262.	1.7743	.15019
-1.6	-8271.	1.7924	.14513

It is shown by Hill (1963) that as α approaches -3.5, the likelihood will increase. This is demonstrated with the series of calculations presented in Table 8.

Table 8. Behavior of the Likelihood as

$-\alpha$ Approaches $x_{min} = 3.5$

α	$-9997[\mu(\alpha)+\ln\hat{\sigma}(\alpha)]$
$-3.5 + 10^{-4}$	-17,095.
$-3.5 + 10^{-50}$	-26,482.
$-3.5 + 10^{-500}$	-35,879.
$-3.5 + 10^{-1000}$	-27,822.
$-3.5 + 10^{-2000}$	-4,819.

Both Harter and Moore (1966) and Hill (1963) recognized this problem. The former simply recommended that the unreasonable parameter estimater giving rise to the large likelihood simply be disallowed. Hill (1963) chose to accomplish this by using a Bayesian type argument.

D. Likelihood when Grouping Errors are Recognized

When it is recognized that all observations are in fact discrete, with a grouping error specified by the methods used to obtain the data, one is immediately led to the definition of the likelihood given in (7.5). As mentioned previously this function is bounded for all values of the parameters. Also, it is continuous and the limiting value is zero when any of the parameters approaches the limits of its range. Consequently, it is known that the function can be maximized, though it is difficult to manipulate, and it is impossible to find an analytic solution for the maximum likelihood estimates, i.e., the values which maximize the function. However, estimates can be obtained by examining the surface with numerical techniques.

The function defined in (7.5) was examined in detail for the data given in Table 6. If f_i represents the frequency with which i is observed, then (7.5) becomes

$$L(\alpha,\mu,\sigma) = K \prod_{i=3}^{27} \left\{ \frac{1}{\sqrt{2\pi}\sigma} \int_{\max(i-\frac{1}{2},-\alpha)}^{i+\frac{1}{2}} \exp -\left\{ \tfrac{1}{2} \frac{\ln(x+\alpha)-\mu}{\sigma} \right\}^2 \frac{dx}{x+\alpha} \right\}^{f_i} \tag{7.8}$$

for $-\alpha < 3.5$, $\sigma > 0$ and $-\infty < \mu < \infty$

and zero otherwise.

K is a constant term which does not depend on the three parameters. Because of the complexity of (7.8) the quantity

$$L^{*} = \ln L(\alpha,\mu,\sigma) - \ln K \tag{7.9}$$

was maximized for selected values of α by using an iterative technique developed by Powell (1964). The results of these computations are shown in Table 9.

An examination of the results reveals that the likelihood attains its maximum at $\hat{\alpha} = -3.0$, $\hat{\sigma} = .498$ and $\hat{\mu} = 1.499$. These values are extremely close to the known true values, $\alpha = -3.$, $\sigma = .5$ and $\mu = 1.5$. The very rapid increase in the value of the likelihood as α advanced from -3.1 to $-3.$ caused some concern and was investigated more thoroughly. The logarithm of the likelihood function (minus a constant

Table 9. The Likelihood Function for Selected Values of α when Grouping Errors are taken into Account

α	$\hat{\sigma}(\alpha)$	$\hat{\mu}(\alpha)$	L^*
-3.999	.661	1.204	-22713.
-3.995	.660	1.205	-22711.
-3.990	.658	1.206	-22710.
-3.970	.656	1.216	-22703.
-3.950	.650	1.220	-22696.
-3.930	.645	1.228	-22690.
-3.910	.640	1.234	-22684.
-3.9	.637	1.239	-22682.
-3.8	.617	1.273	-22657.
-3.2	.601	1.307	-22638.
-3.6	.581	1.336	-22623.
-3.5	.566	1.366	-22612.
-3.4	.548	1.397	-22605.
-3.3	.535	1.423	-22601.
-3.2	.522	1.452	-22599.
-3.1	.509	1.477	-22599.
-3.0	.498	1.499	-22388.
-2.9	.486	1.524	-22389.
-2.8	.478	1.549	-22391.
-2.7	.466	1.571	-22394.
-2.6	.457	1.594	-22399.
-2.5	.446	1.616	-22405.
-2.4	.438	1.639	-22411.
-2.3	.430	1.660	-22418.
-2.2	.421	1.680	-22426.
-2.1	.413	1.699	-22434.
-2.0	.406	1.719	-22442.
-1.9	.397	1.738	-22451.
-1.8	.392	1.758	-22460.
-1.7	.385	1.774	-22470.
-1.6	.379	1.793	-22479.
-1.5	.372	1.812	-22489.
-1.4	.367	1.829	-22498.
-1.3	.362	1.845	-22508.
-1.2	.355	1.862	-22518.
-1.5	.350	1.878	-22528.

Table 10. Logarithm of the Likelihood Minus a Constant Term for Selected Parameter Values

σ	μ	α=-3.10	α=-3.05	α=-3.00	α=-2.95
	.498	-22604.3	-22601.1	-22393.0	-22403.4
	.500	.22603.1	-22600.7	-22392.9	-22404.1
	.502	-22602.1	-22600.6	-22392.8	-22405.1
1.485	.504	-22601.3	-22600.7	-22393.3	-22406.2
	.506	-22601.1	-22601.2	-22394.0	-22407.6
	.508	-22600.8	-22601.8	-22394.5	-22409.4
	.510	-22600.9	-22602.9	-22396.4	-22411.3
	.498	-22606.4	-22600.8	-22390.5	-22398.7
	.500	-22605.2	-22600.4	-22390.6	-22399.3
	.502	-22604.2	-22600.3	-22390.8	-22400.3
1.490	.504	-22603.3	-22600.3	-22391.2	-22401.6
	.506	-22603.1	-22601.1	-22391.9	-22403.1
	.508	-22602.9	-22601.6	-22393.0	-22404.9
	.510	-22603.1	-22602.7	-22394.1	-22407.0
	.498	-22609.5	-22601.5	-22389.4	-22395.0
	.500	-22608.2	-22601.1	-22389.1	-22395.7
	.502	-22607.2	-22600.8	-22389.4	-22396.7
1.495	.504	-22606.5	-22601.2	-22389.9	-22398.0
	.506	-22606.1	-22601.5	-22390.6	-22399.5
	.508	-22606.0	-22602.2	-22391.6	-22401.4
	.510	-22606.0	-22603.4	-22393.1	-22403.4
	.498	-22613.6	-22603.0	-22388.7	-22392.1
	.500	-22612.2	-22602.6	-22388.7	-22392.7
	.502	-22611.2	-22602.5	-22389.0	-22393.9
1.500	.504	-22610.6	-22602.7	-22389.6	-22395.3
	.506	-22610.2	-22603.1	-22390.4	-22396.9
	.508	-22609.7	-22604.1	-22391.4	-22398.6
	.510	-22610.1	-22604.9	-22392.7	-22400.8

term) was evaluated at a grid of points. These calculations are summarized in Table 10. The extremely steep drop in the likelihood between $\alpha = -3.05$ and $\alpha = -3.00$ is again evident in this table. The maximum attained in Table 9 does not occur on this table because that point did not happen to coincide with one of the grid points. The maximum on the grid is -22388.7 and corresponds to the parameter values $\alpha = -3.00$, $\sigma = 1.500$ and $\mu = .500$.

An examination of the likelihood which made allowance for the grouping error in the data used by Hill (1963) highlighted the major difficulty with this method. A thorough examination of the surface, using a computer which carried 19 binary bits per word (equivalent to slightly more than 5 decimal digits) indicated that the maximum likelihood estimates were $\hat{\alpha} = -3.22$, $\hat{\mu} = 1.76$ and $\hat{\sigma}^2 = .074$. The corresponding estimates obtained by Hill were -1.59, 2.01 and .042 (-4- Hill's $\hat{\gamma} = -4-(-2.41) = -1.59$). A careful evaluation of the likelihood function incorporating the grouping error, using 15 place tables of the normal integral and logarithms indicated that Hill's solution actually lead to a slightly larger likelihood. The difficulty was caused by the one out-lying observation in the data. It appears that in order to use this method successfully one must have access to a computer which carries a large number of digits of accuracy, and use very accurate routines to compute integrals and logarithms. It is also interesting to note that both sets of parameter estimates obtained above lead to large χ^2 values suggesting either that the three-parameter lognormal does not fit the data, or that the correct parameter estimates have not been obtained.

E. Locating the Maximum of the Likelihood Function

Since powerful procedures and associated computer programs for maximizing (or minimizing) functions of several variables are still quite novel, it was decided to include a discussion of the method used for this problem. The program used is a Fortran II adaptation of one originally written by Powell (1964). One of the features of the method used in this program is that the derivatives of the function are never calculated.

Assume that a function of m parameters is to be maximized. Each iteration of the procedure commences with a search along m linearly independent directions ξ_1 , ξ_2 , ..., ξ_m , starting from an initial approximation to the answer. Initially the directions are the co-ordinate directions, so that the first iteration is identical to the method which changes one parameter at a time. However, each iteration defines a new direction ξ , and the directions for the next iteration are ξ_2 , ξ_3 , ..., ξ_m , . The procedure for choosing the new direction each iteration is such that the conjugate directions are generated. If a quadratic function is being maximized the method is such that after m iterations all the directions are mutually conjugate and the next iteration will yield the exact optimum.

The program, however, is more complex than is indicated above in order to ensure reasonable rates of convergence for more general functions.

One of the modifications is to allow a direction other than ξ_1 to be discarded, so that the new direction will always contain an appreciable component of the direction which is lost. This prevents the procedure from generating nearly dependent directions.

The second modification of the procedure concerns the stopping rule. The basic stopping rule is to assume convergence when an iteration causes each variable to change by less than the accuracy specified by the user. This has been modified however, so that when this criterion is satisfied, say at point A, each variable is increased by 100 times the specified accuracy. The normal procedure for maximization is then carried on until all changes are again less than the specified accuracy. Let the resulting point be B. The maximum on the line connecting A and B is then located. Call this C. Now if all components on the lines joining C with A and B are less than the specified accuracy, the procedure is terminated. If this is not the case, ξ_1 is replaced by the direction specified by A and C. This modification allows the procedure to get away from a region where the function changes very slowly.

The method used to compute the normal integrals is described in Hastings (1955). The approximation, as given there has an absolute error $< 10^{-6}$ to a range of more than six standard deviations from the mean. However, this would be reduced somewhat by the limitations of the computer. The logarithms were also computed using an approximation from Hastings (1955). In order to prevent difficulties in the computer, allowance had to be made for the occurrance of zero and possibly even negative (very small in absolute value) probabilities due to inaccuracies in the approximation and truncation errors. The procedure adopted was to test for a zero or negative probability value before computing the logarithm. If either of these occurred, the probability was set equal to $(.1)^{10}$.

F. Variances and Covariances of the Maximum Likelihood Estimates

Clearly it is not possible to obtain exact expressions for the variances and covariances of the estimates of the three parameters in (7.5). However, it is well known that for large samples the maximum likelihood estimates $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\alpha}$ will be asymptotically multivariate normal with means μ , σ and α and variance covariance matrix equal to the negative of the inverse of the expected value of the matrix of partial derivatives

$$\begin{pmatrix} \frac{\partial^2 \ln L}{\partial\mu^2} & \frac{\partial^2 \ln L}{\partial\mu\partial\sigma} & \frac{\partial^2 \ln L}{\partial\mu\partial\alpha} \\ \frac{\partial^2 \ln L}{\partial\mu\partial\sigma} & \frac{\partial^2 \ln L}{\partial\sigma^2} & \frac{\partial^2 \ln L}{\partial\alpha\partial\sigma} \\ \frac{\partial^2 \ln L}{\partial\mu\partial\alpha} & \frac{\partial^2 \ln L}{\partial\alpha\partial\sigma} & \frac{\partial^2 \ln L}{\partial\alpha^2} \end{pmatrix} . \qquad (7.10)$$

If the constant K is ignored, (7.5) can be rewritten as

$$L^*(\alpha,\mu,\sigma) = \prod_i \left\{ \int_{\frac{\ln(\max((i-\frac{1}{2})\Delta,-\alpha)+\alpha)-\mu}{\sigma}}^{\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp - \frac{x^2}{2}\, dx \right\}^{f_i} \qquad (7.11)$$

where $\sigma > 0$, $-\infty < \mu < \infty$, $-\alpha < (i'+\frac{1}{2})\Delta$ and i' is the minimum value of i for which $f_i \neq 0$. Otherwise $L^*(\alpha,\mu,\sigma)$ is defined as zero. f_i is the number of times the value Δi has been observed in the sample.

If one defines

$$G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp - y^2/2\, dy$$

then

$$\ln L^* = \sum_i f_i \ln\left[G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right) - G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)\right] . \qquad (7.12)$$

Since L^* defined by (7.10) and L defined by (7.5) differ by a constant which does not depend on any parameters, the partial derivatives of the logarithms of the two terms will be equal.

The first partial derivatives of $\ln L^*$ are:

$$\frac{\partial \ln L^*}{\partial \mu} \tag{7.13}$$

$$= \sum_i f_i \left\{ \frac{\dfrac{\partial G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right)}{\partial \mu} - \dfrac{\partial G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)}{\partial \mu}}{G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right) - G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)} \right\}$$

where

$$\frac{\partial G\left(\frac{\ln(x)-\mu}{\sigma}\right)}{\partial \mu} = - \frac{-1}{\sqrt{2\pi}\sigma} \exp - \tfrac{1}{2}\left\{\frac{\ln(x)-\mu}{\sigma}\right\}^2 .$$

$$\frac{\partial \ln L^*}{\partial \sigma} \tag{7.14}$$

$$= \sum_i f_i \left\{ \frac{\dfrac{\partial G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right)}{\partial \sigma} - \dfrac{\partial G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)}{\partial \sigma}}{G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right) - G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)} \right\}$$

where

$$\frac{\partial G\left(\frac{\ln(x)-\mu}{\sigma}\right)}{\partial \sigma} = \frac{\ln(x)-\mu}{\sqrt{2\pi}\sigma^2} \exp - \tfrac{1}{2}\left\{\frac{\ln(x)-\mu}{\sigma}\right\}^2 .$$

$$\frac{\partial \ln L^*}{\partial \alpha} \tag{7.15}$$

$$= \sum_i f_i \left\{ \frac{\dfrac{\partial G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right)}{\partial \alpha} - \dfrac{\partial G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)}{\partial \alpha}}{G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right) - G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,-\alpha)-\mu}{\sigma}\right)} \right\}$$

where

$$\frac{\partial G\left(\frac{\ln(x)-\mu}{\sigma}\right)}{\partial \alpha} = \frac{1}{\sqrt{2\pi}\ ((i\pm\frac{1}{2})\Delta+\alpha)\sigma} \exp - \tfrac{1}{2}\left\{\frac{\ln(x)-\mu}{\sigma}\right\}^2$$

for $(i\pm\frac{1}{2})\Delta > -\alpha$. For the special case where $(i-\frac{1}{2})\Delta < -\alpha$,

$$G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)$$

$$= G\left(\frac{\ln(-\alpha+\alpha)-\mu}{\sigma}\right)$$

$$= 0 \ ,$$

and $\dfrac{\partial G\left(\frac{\ln(0)-\mu}{\sigma}\right)}{\partial \alpha} = 0$.

The six second partial derivatives of $\ln L^*$ are:

$$\frac{\partial^2 \ln L^*}{\partial\mu^2} \tag{7.16}$$

$$= \sum_i f_i \left\{ \frac{\dfrac{\partial^2 G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial\mu^2} - \dfrac{\partial^2 G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})}{\partial\mu^2}}{G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})} \right\}$$

$$- \sum_i f_i \left\{ \frac{\dfrac{\partial G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial\mu} - \dfrac{\partial G(\frac{\ln(\max((i-\frac{1}{2})\Delta,-\alpha)+\alpha)-\mu}{\sigma})}{\partial\mu}}{G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)-\mu}{\sigma})} \right\}^2$$

where

$$\frac{\partial^2 G(\frac{\ln(x)-\mu}{\sigma})}{\partial\mu^2} = \frac{\ln(x)-\mu}{\sigma^2} \cdot \partial G(\frac{\ln(x)-\mu}{\partial\sigma}) .$$

$$\frac{\partial^2 \ln L^*}{\partial\sigma^2} \tag{7.17}$$

$$= \sum_i f_i \left\{ \frac{\dfrac{\partial^2 G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial\sigma^2} - \dfrac{\partial^2 G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})}{\partial\sigma^2}}{G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma} - G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)-\mu}{\sigma})} \right\}$$

$$- \sum_i f_i \left\{ \frac{ \dfrac{\partial G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial \sigma} - \dfrac{\partial G(\frac{\ln(\max((i-\frac{1}{2}) \ , -\alpha)+\alpha)-\mu}{\sigma})}{\partial \sigma} }{ G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta, -\alpha)-\mu}{\sigma}) } \right\}^2$$

where

$$\frac{\partial^2 G(\frac{\ln(x)-\mu}{\sigma})}{\partial \sigma^2} = \frac{\partial G(\frac{\ln(x)-\mu}{\sigma})}{\partial \sigma} \cdot \frac{\ln(x)-\mu - \partial\sigma^2}{\sigma^3} .$$

$$\frac{\partial^2 \ln L^*}{\partial \alpha^2} \tag{7.18}$$

$$= \sum_i f_i \left\{ \frac{ \dfrac{\partial^2 G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial \alpha^2} - \dfrac{\partial^2 G(\frac{\ln(\max((i-\frac{1}{2})\Delta, -\alpha)+\alpha)-\mu}{\sigma})}{\partial \alpha^2} }{ G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta, -\alpha)+\alpha)-\mu}{\sigma}) } \right\}$$

$$- \sum_i f_i \left\{ \frac{ \dfrac{\partial G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial \alpha} - \dfrac{\partial G(\frac{\ln(\max((i-\frac{1}{2})\Delta,-\alpha)-\mu}{\sigma})}{\partial \alpha} }{ G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta, -\alpha)+\alpha)-\mu}{\sigma}) } \right\}^2$$

where

$$\frac{\partial^2 G(\frac{\ln(x)-\mu}{\sigma})}{\partial \alpha^2} = - \frac{\partial G}{\partial \alpha} \frac{[\ln((i\pm\frac{1}{2})\Delta+\alpha)-\mu]^2-\sigma^2}{\sigma^2}$$

for $(i\pm\frac{1}{2})\Delta > -\alpha$. For the special case where $(i-\frac{1}{2})\Delta < -\alpha$, G is a constant, equal to zero and hence $\frac{\partial^2 G}{\partial \alpha^2} = 0$.

$$\frac{\partial^2 \ln L^*}{\partial\mu\partial\sigma} \tag{7.19}$$

$$= \sum_i f_i \left\{ \frac{\dfrac{\partial^2 G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial\mu\partial\sigma} - \dfrac{\partial^2 G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})}{\partial\mu\partial\sigma}}{G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta,-\alpha)+\alpha)-\mu}{\sigma})} \right\}$$

$$- \sum_i f_i \left\{ \frac{\dfrac{\partial G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial\mu} \quad \dfrac{\partial G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})}{\partial\mu}}{G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})} \right\} \cdot$$

$$\left\{ \frac{\dfrac{\partial G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial\sigma} - \dfrac{\partial G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)-\mu}{\sigma})}{\partial\sigma}}{G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)-\mu}{\sigma})} \right\}$$

where

$$\frac{\partial^2 G(\frac{\ln(x)-\mu}{\sigma})}{\partial\mu\partial\sigma}$$

$$= \frac{\partial G(\frac{\ln(x)-\mu}{\sigma})}{\partial\mu} \cdot \left\{\frac{(\ln(x)-\mu)^2-\sigma^2}{\sigma^3}\right\} .$$

$$\frac{\partial^2 \ln L^*}{\partial\mu\partial\alpha} \tag{7.20}$$

$$= \sum_i f_i \left\{ \frac{\dfrac{\partial^2 G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right)}{\partial\mu\partial\alpha} - \dfrac{\partial^2 G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)-\mu}{\sigma}\right)}{\partial\mu\partial\alpha}}{G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right) - G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\partial\mu\partial\alpha}\right)} \right\}$$

$$- \sum_i f_i \left\{ \frac{\dfrac{\partial G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right)}{\partial\alpha} - \dfrac{\partial G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)-\mu}{\sigma}\right)}{\partial\alpha}}{G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right) - G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)} \right\} \cdot$$

$$\left\{ \frac{\dfrac{\partial G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right)}{\partial\mu} - \dfrac{\partial G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)}{\partial\mu}}{G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right) - G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)-\mu}{\sigma}\right)} \right\}$$

where

$$\frac{\partial^2 G\left(\frac{\ln(x)-\mu}{\sigma}\right)}{\partial\mu\partial\alpha} = \frac{\partial G\left(\frac{\ln(x)-\mu}{\sigma}\right)}{\partial\alpha} \cdot \left\{\frac{\ln(x)-\mu}{\sigma}\right\} .$$

$$\frac{\partial^2 \ln L^*}{\partial\sigma\partial\alpha} \tag{7.21}$$

$$= \sum_i f_i \left\{ \frac{\dfrac{\partial^2 G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right)}{\partial\sigma\partial\alpha} - \dfrac{\partial^2 G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)}{\partial\sigma\partial\alpha}}{G\left(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}\right) - G\left(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma}\right)} \right\}$$

$$- \sum_i f_i \left\{ \frac{\dfrac{\partial G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial\sigma} \quad \dfrac{\partial G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})}{\partial\sigma}}{G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})} \right\} \cdot$$

$$\left\{ \frac{\dfrac{\partial G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma})}{\partial\alpha} - \dfrac{\partial G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})}{\partial\alpha}}{G(\frac{\ln((i+\frac{1}{2})\Delta+\alpha)-\mu}{\sigma}) - G(\frac{\ln(\max((i-\frac{1}{2})\Delta,\ -\alpha)+\alpha)-\mu}{\sigma})} \right\}$$

where

$$\frac{\partial^2 G(\frac{\ln(x)-\mu}{\sigma})}{\partial\sigma\partial\alpha} = - \frac{\partial G(\frac{\ln(x)-\mu}{\sigma})}{\partial\alpha} \cdot \{\frac{\ln(x)-\mu+\sigma}{\sigma^2}\} .$$

Obviously the expected values of expressions (7.16) to (7.21) cannot be obtained for arbitrary values of the parameters. However, consistent estimates of these quantities can be obtained by substituting the maximum likelihood estimates for the parameters and evaluating the indicated sums for the sample data.

G. Numerical Results

In order to provide some insight into the behavior of the variances and covariances of the maximum likelihood estimates for large samples, the expected value of the six second order partials, given by equations (7.16) to (7.21) were evaluated numerically for $\mu = 1.5$, $\sigma = .5$ and $\alpha = 3$. These calculations were repeated with a series of six different grouping errors, i.e., Δ values. The expected values were obtained by substituting the assumed parameter values in the functions, replacing f_i by the probability that an observation will fall in the interval $(i-\frac{1}{2})\Delta < x < (i+\frac{1}{2})\Delta$ and performing the indicated summation. Of course, it is assumed that any of these numbers would be recorded as equal to Δi. Also the calculations were set up so that the point x = 3. corresponded exactly to the lower endpoint of a grouping interval. It is to be remembered that if one thinks in terms of a sample of size n, then the calculations need to be increased by a factor of n. These computations are summarized in Table 11.

Table 11. Expected value of Second Order Partial Derivatives of the Three Parameter Lognormal Distribution

Δ	$\frac{\partial^2 \ln L^*}{\partial \mu^2}$	$\frac{\partial^2 \ln L^*}{\partial \sigma^2}$	$\frac{\partial^2 \ln L^*}{\partial \alpha^2}$	$\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma}$	$\frac{\partial^2 \ln L^*}{\partial \mu \partial \alpha}$	$\frac{\partial^2 \ln L^*}{\partial \sigma \partial \alpha}$
3.	-3.45700	-5.45209	-.23368	.56876	.78307	.40255
2.	-3.68055	-6.54644	-.30840	.38733	.87537	.66431
1.	-3.90088	-7.27573	-.36202	.14611	.95765	.84352
.5	-3.97330	-7.72033	-.39392	.01898	.99567	.95846
.3	-3.99002	-7.83718	-.40379	-.01382	1.00535	.99123
.2	-3.99538	-7.87629	-.40733	-.02500	1.00859	1.00266

For a fixed Δ the variance-covariance matrix of the limiting multivariate normal distribution of the three maximum likelihood estimates can be obtained by inverting the negative of the matrix of expected second order partials and dividing by n, the sample size.

For Δ = 3. the resulting matrix is

$$1/n \begin{pmatrix} 16.630,26 & 6.701,87 & 67.272,26 \\ 6.701,87 & 2.910,95 & 27.472,37 \\ 67.272,26 & 27.472,37 & 277.033,50 \end{pmatrix} .$$

For Δ = 2. the matrix is

$$1/n \begin{pmatrix} 5.378,66 & 2.389,91 & 20.415,08 \\ 2.389,91 & 1.257,39 & 9.492,14 \\ 20.415,08 & 9.492,14 & 81.636,47 \end{pmatrix} ,$$

and for Δ = 1. the matrix is

$$1/n \begin{pmatrix} 3.297,85 & 1.476,47 & 12.163,97 \\ 1.476,47 & .849,34 & 5.884,68 \\ 12.163,97 & 5.884,68 & 48.650,97 \end{pmatrix} .$$

These three matrices indicate the trend in the variances and covariances as Δ becomes smaller. As one would expect, all variances are becoming smaller, and especially the variance of $\hat{\alpha}$. A striking feature of this series of matrices is the extremely large variances of $\hat{\mu}$ and $\hat{\sigma}$. The trend in the elements of the matrix continues, as indicated for the next three Δ values.

For Δ = .5 the matrix is

$$1/n \begin{pmatrix} 2.855,03 & 1.293,70 & 10.364,16 \\ 1.293,70 & .771,80 & 5.147,87 \\ 10.364,16 & 5.147,87 & 41.260,67 \end{pmatrix} .$$

For Δ = .3 the matrix is

$$1/n \begin{pmatrix} 2.685,21 & 1.219,46 & 9.679,07 \\ 1.219,46 & .738,86 & 4.849,92 \\ 9.679,07 & 4.849,92 & 38.480,74 \end{pmatrix} .$$

Finally for Δ = .2 the matrix is

$$1/n \begin{pmatrix} 2.623,84 & 1.192,38 & 9.432,03 \\ 1.192,38 & .726,78 & 4.741,48 \\ 9.432,03 & 4.741,48 & 37.481,24 \end{pmatrix} .$$

As pointed out previously, a striking feature of these matrices is the large variances of $\hat{\mu}$ and $\hat{\sigma}$. For comparison consider the asymptotic variance-covariance matrix of the maximum likelihood estimates of $\hat{\mu}$ and $\hat{\sigma}$ for the usual two parameter lognormal distribution when the grouping error is ignored. In this case the variance-covariance matrix is

$$1/n \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix} .$$

For σ = .5 this matrix becomes

$$1/n \begin{pmatrix} .25 & 0 \\ 0 & .125 \end{pmatrix} .$$

Now to show that the inflated variances for $\hat{\mu}$ and $\hat{\sigma}$ were due to the α parameter, consider what happens when it is assumed that α is known and the grouping error is introduced. The following series of variance covariance matrices results:

For Δ = 3.

$$1/n \begin{pmatrix} .294,32 & .030,70 \\ .030,70 & .186,62 \end{pmatrix} .$$

For Δ = 2.

$$1/n \begin{pmatrix} .273,40 & .016,18 \\ .016,17 & .153,71 \end{pmatrix} .$$

For Δ = 1.

$$1/n \begin{pmatrix} .256,55 & .005,15 \\ .005,15 & .137,55 \end{pmatrix} .$$

For Δ = .5

$$1/n \begin{pmatrix} .251,68 & .000,62 \\ .000,62 & .129,53 \end{pmatrix} .$$

For $\Delta = .3$

$$1/n \begin{pmatrix} .250,63 & -.000,44 \\ -.000,44 & .127,60 \end{pmatrix} .$$

For $\Delta = .2$

$$1/n \begin{pmatrix} .250,29 & -.000,79 \\ -.000,79 & .126,97 \end{pmatrix} .$$

This series of matrices indicates the trend towards the matrix for the case where there is no grouping error. Also the elements are of the correct order of magnitude.

Another interesting feature of the three parameter case is the large correlation induced between $\hat{\mu}$ and $\hat{\sigma}$. For the six values of Δ ranging from $\Delta = 3$ to $\Delta = .2$ the correlation takes on the values .963, .919, .882, .872, .866 and finally .863. By contrast, this correlation has the value of .130 when α is known and $\Delta = 3$.

The above series of calculations demonstrates clearly the effect of the grouping error in causing an increase in the variance of the estimates. This is particularly marked when the threshold parameter α is unknown. A sample from the distribution considered above will have a mean approximately equal to 8.1 and a variance of 7.3. Hence the grouping error $\Delta = 3.$ which amounts to slightly more than one standard deviation has destroyed almost all the information about α present in the sample. In fact even with $\Delta = .2$, i.e., a grouping interval of less than one tenth of one standard error, the variance of α is still appreciable. Alternatively the effect of the grouping error appears to be much less drastic when α is known.

VIII. TRANSFORMATIONS WHICH MAXIMIZE THE CORRELATION BETWEEN RESIDUALS AND THE NORMAL ORDER STATISTICS

A. Introduction

When dealing with relatively simple problems, such as estimating a location parameter, and it is desirable to eliminate the effects of gross outliers, a natural practice is to remove an equal number of the largest and smallest observations and then proceed as if the trimmed sample were the whole sample. Tukey (1962) discusses a modification of this which he calls Winsorizing, which is simply to replace any outlying value by the nearest number recorded for an observation which is not seriously suspect. While the first procedure is simply a rejection technique, the latter can be thought of as a monotone transformation.

The process of Winsorizing does not extend directly to data sets having a more complex structure than the simple random sample. The problem is that while there is a simple relationship between the ordering of the observations and the errors involved, this is not true for more complex data structures. For example, in a two-way array, one observation may be larger than another, either because of the effects of a large positive error, or because of larger row and column effects.

Tukey (1962) describes a method of analysis for the two-way array in which the effect that particular deviatons are permitted to have on the final estimates are modified according to the size of the deviation. Large deviations have negligible effects and small deviations are permitted larger effects. The major problem in this method is to decide when a value deviates too much from the main body of the data. The procedure adopted is to fit row and column means, calculate residuals of apparent deviations, order them, plot them against typical values for normal order statistics, draw a straight line through the result and then use the distance which values lie off the line as a measure of the amount of deviation.

The technique to be developed in this chapter is similar to the above in that ordered residuals are compared with typical values for normal order statistics. The object will be to find a monotone transformation of the data which maximizes the correlation between the ordered residuals and the order statistics. The motivation is that in the course of an experiment, a scale for measuring the phenomenon of interest must be chosen. If this scale is chosen correctly, the residuals remaining after fitting a linear model will be normally distributed. This, however, will not be the case if an incorrect scale was selected. Consequently, a criterion to use for finding a transformation is that it modify the scale of measurement so that the residuals are normally distributed. Ideally one should compare the ordered residuals to the normal order statistics from a correlated sample. However, these order statistics are not available. Also, if the ratio of error degrees of freedom to degrees of freedom for effects is large, the correlation between residuals will tend to become small on the average, and the order statistics from a simple sample used as an approximation.

B. Description of the Method

Let $\{y_i\}$ be the n observed numbers arranged so that $y_i \leq y_j$ if $i < j$. Let $\{\eta_i\}$ be the set of n variables obtained from the observed numbers by means of a monotone transformation. It follows that $\eta_i \leq \eta_j$ if $i < j$. The aim will be to choose a transformation which results in the $\{\eta_i\}$ satisfying the ideal statistical conditions, i.e., additive model and normal independent errors.

The procedure will be illustrated in this chapter by considering a function of the form

$$\eta_i = y_i + by_i^2 \tag{8.1}$$

where the choice of b depends on the $\{y_i\}$ and is chosen so that η_i is always less than η_j if y_i is less than y_j . b is chosen subject to the above restrictions so as to maximize the correlation between the ordered residuals and the normal order statistics for a sample of size n . The method can be extended immediately to a function of the type

$$\eta_i = y_i + by_i^2 + cy_i^3 \tag{8.2}$$

or even higher order polynomials, with restrictions on the parameters ensuring that the transformation is monotone for the observed set of data. However, only the simpler case given in (8.1) will be discussed here.

For the function given by (8.1) the suggested procedure for finding b is to compute the correlation between the ordered residuals and the n normal order statistics for a series of choices for b and then interpolating to determine the optimum. The calculations are simple and straight-forward. For a given b value, use (8.1) to compute the $\{\eta_i\}$. Perform the analysis on these transformed values and obtain the set of residuals $\{z_i\}$. From these obtain the ordered set $\{z_i^*\}$ where $z_i^* \leq z_j^*$ if $i < j$. Then compute the correlation between the $\{z_i^*\}$ and the normal order statistics for a sample of size n . After performing these calculations for a series of b values, the optimum can be deduced. Now let b^* be the optimum value. Compute the set $\{\eta_i\}$ corresponding to b^* and plot the values against the observed $\{y_i\}$. This graph will then suggest the appropriate form of the transformation. This step can be eliminated if one is willing to accept (8.1) as the transformation. However, the plot outlined above will often suggest a more conventional transformation, such as the square root or the logarithm.

C. Numerical Examples

In order to demonstrate the procedure outlined above a set of data was constructed and analyzed much as an experimenter would. The basic data consisted of 16 observations for a 2^4 experiment in which $\mu = 10$, A = 6, B = 4, C = 4, AB = 2 and all other effects and interactions were zero. Independent normal errors with mean zero and variance 1 were attached to the values. The resulting synthetic data are as follows:

(1)	=	4.33	c	=	7.56	d	=	3.86	cd	=	7.58
a	=	8.15	ac	=	11.00	ad	=	9.08	acd	=	12.59
b	=	5.67	bc	=	10.60	bd	=	5.50	bcd	=	10.08
ab	=	15.05	abc	=	17.18	abd	=	14.37	abcd	=	17.97

An analysis of variance on this 'true' data using the three and four factor interactions as errors indicated that A, B, AB and C were all significant at the 1% level and AD is just significant at the 5% level. This is essentially what one would expect.

Now assume that the experimenter read his numbers on an incorrect scale and observed:

(1)	=	19	c	=	57	d	=	15	cd	=	57
a	=	66	ac	=	121	ad	=	82	acd	=	158
b	=	32	bc	=	113	bd	=	30	bcd	=	102
ab	=	227	abc	=	123	abd	=	206	abcd	=	323

These numbers were obtained from the synthetic data by squaring the numbers. An analysis of variance of this data using the three and four factor interactions for error yields a test for A which is significant at the 1% level, a test for B significant at the 5% level and non-significant values for the other effects and interactions. As far as the experimenter is concerned there has been a loss of information due to the use of the incorrect scale even if he is not aware of this fact.

From here on all analyses are based on the assumptions that the three and four factor interactions measure error. This does not agree with the true state of affairs which we know exists since the data were artificially constructed. However, this should tend to make the simulation of a real world situation more realistic.

The order statistics for a sample of 16 from the normal distribution are -1.7, -1.3, -1.0, -.8, -.6, -.4, -.2, -.1, .1, .2, .4, .6, .8, 1.0, 1.3 and 1.7.

The correlations between the residuals obtained from the analysis of the values $y_i + by_i^2$ for i = 1, ..., 16 assuming that three and four factors interactions were error contrasts and the normal order statistics were calculated for various values of b . The following array of values was obtained:

b	Correlation squared × 13.58
-.0018	13.10
-.0016	13.26
-.0014	13.25
-.0010	12.77
0	12.01
.0010	12.13
.0024	12.48
.003	12.56
.004	12.71
.005	12.80
.006	12.87

As the b value was increased the value of the square of the correlation coefficient approached 12.87/13.58. Consequently the maximum value is obtained at b = -.0016. The final transformed values are given in the following table:

(1)	=	18.36	c	=	51.88	d	=	14.68	cd	=	51.88
a	=	58.96	ac	=	97.64	ad	=	71.28	acd	=	118.00
b	=	30.40	bc	=	92.52	bd	=	28.56	bcd	=	85.36
ab	=	144.60	abc	=	98.84	abc	=	138.16	abcd	=	156.12

Figure 3 shows the final transformed values plotted against the original values. This graph strongly suggests that a square root transformation be applied to the data. From the manner in which the data were constructed it is clear that this yields numbers which satisfy the ideal statistical conditions.

For a second example the same basic set of numbers was used again. However, this time it was assumed that the experimenter had recorded the positive square roots of the correct values.

These 'observed' values were:

(1) = 21	c = 28	d = 20	cd = 28
a = 28	ac = 33	ad = 30	acd = 36
b = 24	bc = 33	bd = 24	bcd = 32
ab = 39	abc = 42	abd = 38	abcd = 42

These values were obtained from the original set by computing the square root and multiplying by 10. The analysis was again performed in the same manner as the above example. The series of values given in Table 12 was obtained.

The value of the squared coefficient of correlation approaches 13.236/13.56 as b increases. Table 12 indicates a relative maximum at $b = \infty$. This is equivalent to a transformation which squares all values. This is in perfect agreement with the way in which the data was constructed. However, there is a maximum at b = -.034. The array of transformed variables obtained with the transformation $\eta_i = y_i - .034\, y_i^2$ is as follows:

(1) = 6.01	c = 1.34	d = 6.4	cd = 1.34
a = 1.34	ac = -4.03	ad = -.6	acd = -8.06
b = 4.42	bc = -4.03	bd = 4.42	bcd = -2.82
ab = -12.71	abc = -17.98	abd = -11.10	abcd = -17.98

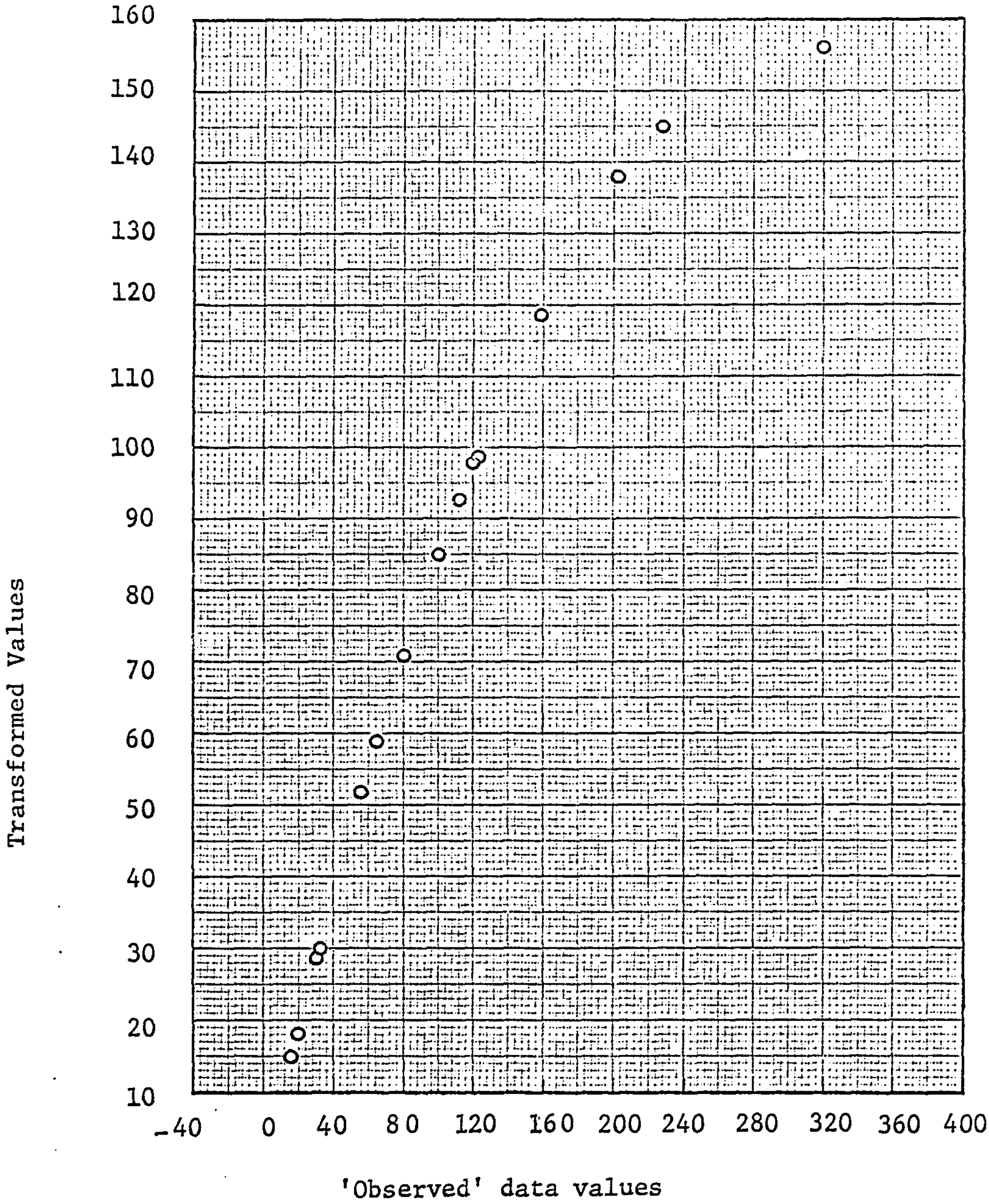

Figure 3. Results of Computations with Data for which the Square Root Transformation is Correct

Table 12. Correlation Between Ordered Residuals and Normal Order Statistics for Various Transformations

b	Correlation squared × 13.58
-1.000	11.13
-.500	13.25
-.100	13.30
-.050	13.37
-.040	13.386
-.037	13.393
-.036	13.394
-.035	13.395
-.034	13.396
-.033	13.394
-.030	13.383
-.020	12.06
-.010	10.81
0	12.65
.010	12.93
.050	13.14
.100	13.18
1.000	13.230
2.000	13.233
5.000	13.235

A graph of these values against the corresponding 'observed' values is shown in Figure 4. The abscissa consists of the 'observed' values and the ordinate the transformed values. The graph suggests that the values be transformed by squaring them. This again is in perfect agreement with the known facts for the data.

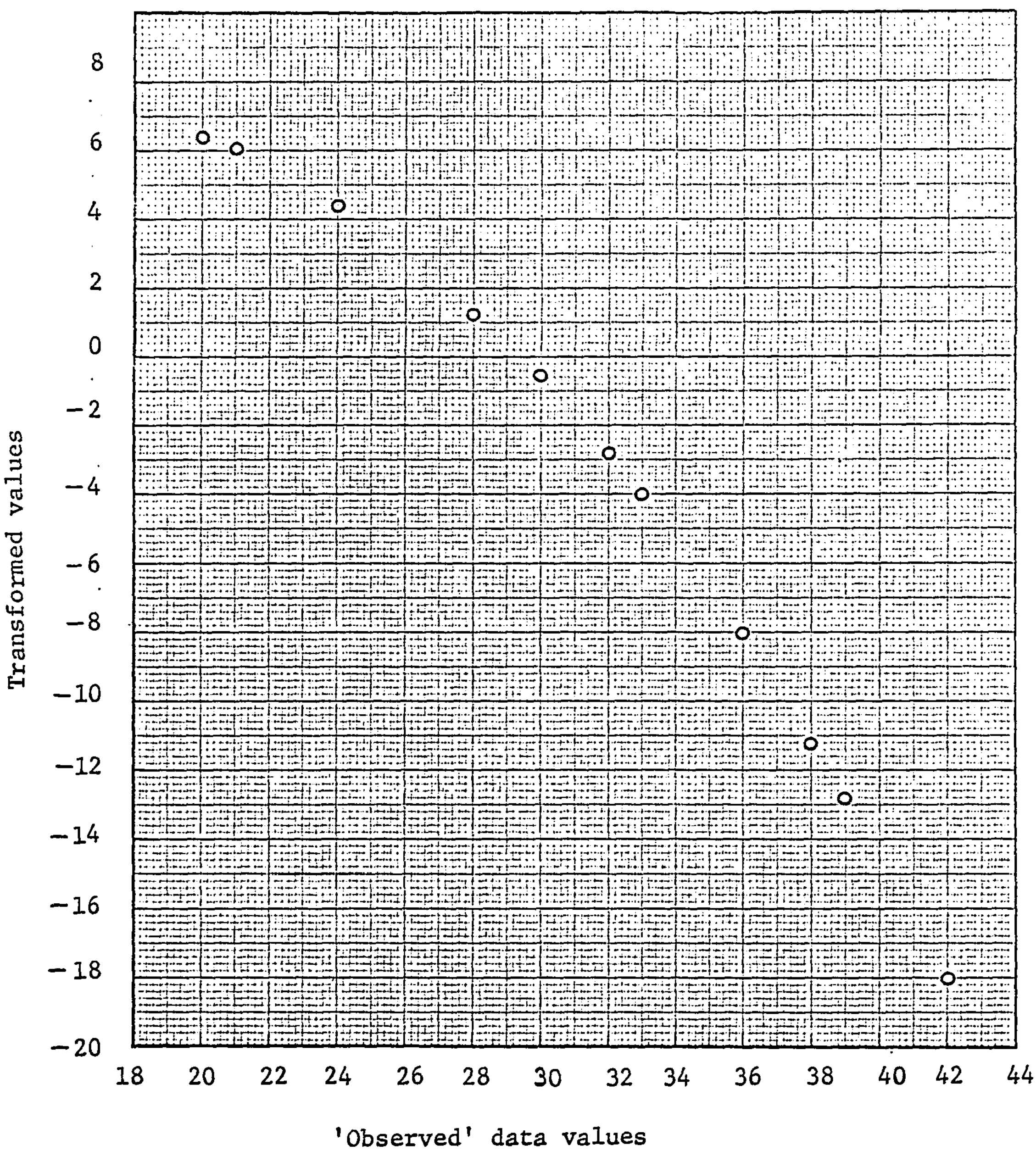

Figure 4. Results of Computations with Data for which the Square is the Correct Transformation

IX. SUMMARY AND CONCLUSIONS

In this thesis a number of methods for examining the residuals after a conventional analysis of variance or least-squares fittings have been explored. A procedure for finding the moments of the residuals has been developed for the case where a linear model holds and the errors are independent random variables form a common distribution. Rules for writing the expected value of products of powers of the residuals in terms of the cumulants of the error distribution are given. If the expected values are to be expressed in terms of moments rather than cumulants, a further substitution is needed. Formulas for the conversion of the first ten cumulants to moments are given in Kendall (1952). Third and fourth order moments have also been obtained for the residuals when the derived linear model with additive block and treatment effects is appropriate.

Properties of two statistics designed to measure departures from ideal statistical conditions (identically distributed, independent normal errors) have been studied under specific non-ideal conditions. One is the first scale-invarient shape coefficient of the residuals measuring skewness. The behavior of this statistic is examined for the case where the errors are independent samples from the Gamma distribution. The second statistic is the first scale-invarient shape coefficient of the residuals measuring kurtosis. It is examined in some detail for the case where the errors are independently drawn from a mixture of two normal distributions with unequal variances. The expected value of the statistic is examined for the effects of changes in the relative sizes of the variances in the two populations and the relative proportions of the two distributions in the mixture.

In Chapter VI a series of tests for non-additivity with more than one degree of freedom is developed for the $r \times c$ table of observations. Procedures for computing the power of some of these tests are developed.

In Chapter VII the possibility of using maximum likelihood estimation procedures for obtaining transformations which would yield a distribution conforming to the ideal statistical conditions were examined. It was found that for the system of power transformations considered in this study, explicit allowance had to be made for the fact that the recorded number must be treated as representing a value in a certain finite interval. If this grouping error was not taken into account, it was shown that the likelihood could be increased without limit by means of suitable choices of values for the parameters. Numerical techniques were required to obtain the maximum likelihood estimates. The effect of varying the grouping interval on the asymptotic variances and covariances of the three parameters in the simple random sample was investigated. The variances were found to increase rather rapidly as the interval was increased. This was especially marked for the estimate of the third or threshold parameter. It was also noted that if this parameter was assumed known, then the asymptotic variances of the estimates of μ and σ were reduced considerably.

In Chapter VIII an iterative procedure for finding a suitable monotone transformation of the data was developed. This procedure was based on a method for finding the transformation which maximized the correlation between the ordered residuals and the normal order statistics. Several numerical examples demonstrating the procedure are given.

X. BIBLIOGRAPHY

Anscombe, F. J. 1955. Contributions to discussion of paper by G. E. P. Box and S. L. Andersen. Royal Statistical Society Journal Series B, 17: 29-30.

Anscombe, F. J. 1960. Rejection of outliers. Technometrics 2: 123-147.

Anscombe, F. J. 1961. Examination of residuals. Berkeley Symposium on Mathematical Statistics and Probability, Fourth, Proceedings 4: 1-36.

Anscombe, F. J. and Tukey, J. W. 1963. The examination and analysis of residuals. Technometrics 5: 141-160.

Box, G. E. P. and Cox, D. R. 1964. An analysis of transformations. Royal Statistical Society Journal 26: 211-252.

Cohen, A. C., Jr. 1951. Estimating parameters of logarithmic-normal distributions by maximum likelihood. American Statistical Association Journal 46: 206-212.

Cramér, H. 1946. Mathematical methods of statistics. Princeton, New Jersey, Princeton University Press.

Dixon, W. J. 1950. Analysis of extreme values. Annals of Mathematical Statistics 21: 488-506.

Dixon, W. J. 1960. Simplified estimation from censored normal samples. Annals of Mathematical Statistics 31: 385-391.

Fisher, R. A. 1921. On the mathematical foundations of theoretical statistics. Royal Society of London Philosophical Transactions Series A, 222: 309-368.

Fisher, R. A. 1958. Statistical methods for research workers. New York, New York, Hafner Publishing Company, Inc.

Ghosh, M. N. and Divakar Sharma. 1963. Power of Tukey's test for non-additivity. Royal Statistical Society Journal Series B, 25: 213-219.

Graybill, F. A. 1961. An introduction to linear statistical models. Vol. 1. New York, New York, McGraw-Hill Book Company, Inc.

Grubbs, F. E. 1950. Sample criteria for testing outlying observations. Annals of Mathematical Statistics 21: 27-58.

Harter, L. H., and Moore, A. H. 1966. Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. American Statistical Association Journal 61: 842-851.

Harvard University Computation Laboratory. 1952. Tables of the error function and its first twenty derivatives. Cambridge, Massachusetts, Harvard University Press.

Hastings, Cecil, Jr. 1955. Approximations for digital computers. Princeton, New Jersey, Princeton University Press.

Hill, Bruce M. 1963. The three-parameter lognormal distribution and bayesian analysis of a point-source epidemic. American Statistical Association Journal 58: 72-84.

Hogben, D. 1963. Some properties of Tukey's test for non-additivity. Unpublished Ph.D. thesis. New Brunswick, New Jersey, Library, Rutgers--The State University.

Huzurbazar, V. S. 1948. The likelihood equation, consistency and the maxima of the likelihood function. Annals of Eugenics 14: 185-200.

Kempthorne, O. 1952. The design and analysis of experiments. New York, New York, John Wiley and Sons, Inc.

Kempthorne, Oscar. 1966. Some aspects of experimental inference. American Statistical Association Journal 61: 11-34.

Kendall, M. G. 1941. Proof of relations connected with the tetrachoric series and its generalization. Biometrika 32: 196-198.

Kendall, M. G. 1952. The advanced theory of statistics. Vol 1. London,England, Charles Griffin and Company, Limited.

Kendall, M. G. and Alan Stuart. 1961. The advanced theory of statistics. Vol. 2. London, England, Charles Griffin and Company, Limited.

Kruskal, J. B. Nonmetric analysis of factorial experiments. Unpublished typewritten manuscript. Murray Hill, New Jersey, Bell Laboratories. ca. 1964.

Moore, P. G. and Tukey, J. W. 1954. Answer to Query 112. Biometrics 10: 562-568.

Müller, J. H. 1930. On the application of continued fractions to the evaluation of certain integrals, with special reference to the incomplete beta-function. Biometrika 22: 284-297.

Powell, M. J. D. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computer Journal 7: 155-162.

Scheffé, H. 1959. The analysis of variance. New York, New York, John Wiley and Sons, Inc.

Tukey, J. W. 1949. One degree of freedom for nonadditivity. Biometrics 5: 232-242.

Tukey, J. W. 1955. Answer to Query 113. Biometrics 11: 111-113.

Tukey, J. W. 1957. On the comparative anatomy of transformations. Annals of Mathematical Statistics 28: 602-632.

Tukey, J. W. 1962. The future of data analysis. Annals of Mathematical Statistics 33: 1-67.

XI. ACKNOWLEDGMENTS

The writer acknowledges the financial support of the National Science Foundation under grant NSF-G14237 during the investigations which led to this thesis. The support of the Research Triangle Institute and Dr. A. L. Finkner who granted the writer free use of the computer during the course of the research and permitted the use of a staff secretary to type the final manuscript.

The understanding and guidance of Professor T. A. Bancroft, Professor A. W. Nordskog and Professor Oscar Kempthorne which facilitated the transfer to the Statistics Department is gratefully acknowledged.

The writer would also like to express his appreciation to Professor Oscar Kempthorne for his attention and advice during the preparation of this thesis and to Mrs. Ruby Monk who undertook the difficult task of typing the final manuscript.

Finally, the writer would like to acknowledge the patience of his wife, Peg, during the many months this thesis was in preparation.

XII. APPENDIX A

A METHOD FOR FINDING THE PROBABILITY OF OBSERVING SPECIFIC RANKINGS OF NORMAL VARIABLES

Let $\{x_i\}$ $i = 1, 2, \ldots, 5$ be independent uncorrelated normal variables with means $\{\mu_i\}$ and variance 1. The probability that on a given trial we will observe the ranking $x_{i_1} > x_{i_2} > x_{i_3} > x_{i_4} > x_{i_5}$ where i_1, i_2, i_3, i_4 and i_5 represent a permutation of the first five integers is equal to the probability that the four variables

$y_1 = \frac{1}{2}(x_{i_1} - x_{i_2})$, $y_2 = \frac{1}{2}(x_{i_2} - x_{i_3})$, $y_3 = \frac{1}{2}(x_{i_3} - x_{i_4})$ and $y_4 = \frac{1}{2}(x_{i_4} - x_{i_5})$ will all be positive. Since the $\{x_i\}$ have continuous distributions, the probability that any two values will be exactly equal is zero and hence can be ignored. The joint distribution of the $\{y_i\}$ is multivariate normal with mean vector

$$\begin{pmatrix} \frac{1}{2}(\mu_{i_1} - \mu_{i_2}) \\ \frac{1}{2}(\mu_{i_2} - \mu_{i_3}) \\ \frac{1}{2}(\mu_{i_3} - \mu_{i_4}) \\ \frac{1}{2}(\mu_{i_4} - \mu_{i_5}) \end{pmatrix}$$

and variance-covariance matrix

$$\begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 1 \end{pmatrix}$$

Let the vector of means be denoted by

$$\begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}.$$

Then the probability that $y_i > 0$ for $i = 1, \ldots 4$ can be written as

$$d = \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \cdot \frac{1}{\pi^2} \frac{1}{\sqrt{7}} e^{-\frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \sigma^{ij} (y_i - \nu_i)(y_j - \nu_j)} dy_1 \, dy_2 \, dy_3 \, dy_4$$

where σ^{ij} represents the ij-th element of the inverse of

$$\begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 1 \end{pmatrix}$$

This can be written more conveniently as

$$d = \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \mathrm{MVN}(\nu_1, \nu_2, \nu_3, \nu_4, \Sigma) \, dy_1 \, dy_2 \, dy_3 \, dy_4 \; .$$

But this is equal to

$$d = \int_{-\nu_1}^{\infty} \int_{-\nu_2}^{\infty} \int_{-\nu_3}^{\infty} \int_{-\nu_4}^{\infty} \mathrm{MVN}(0,\ 0,\ 0,\ 0,\ \Sigma)\ dy_1\ dy_2\ dy_3\ dy_4\ .$$

Now let $\phi(t_1,\ t_2,\ t_3,\ t_4)$ be the corresponding characteristic function defined as

$$\phi(t_1,\ t_2,\ t_3,\ t_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{it_1\ y_1 + it_2\ y_2 + it_3\ y_3 + it_4\ y_4\}$$

$$\cdot\ \mathrm{MVN}(0,\ 0,\ 0,\ 0,\ \Sigma)\ dy_1\ dy_2\ dy_3\ dy_4$$

$$= \exp\{-\ \tfrac{1}{2}(t_1^2 + t_2^2 + t_3^2 + t_4^2 - t_1\ t_2 - t_2\ t_3 - t_3\ t_4)\}$$

$$= \exp\{-\ \tfrac{1}{2}\ t_1^2\} \cdot \exp\{-\ \tfrac{1}{2}\ t_2^2\} \cdot \exp\{-\ \tfrac{1}{2}\ t_3^2\} \cdot \exp\{-\ \tfrac{1}{2}\ t_4^2\}$$

$$\cdot\ \exp\{\tfrac{1}{2}\ t_1\ t_2\} \cdot \exp\{\tfrac{1}{2}\ t_2\ t_3\} \cdot \exp\{\tfrac{1}{2}\ t_3\ t_4\}$$

since all other correlations are zero. Now, since

$$\exp\{\tfrac{1}{2}\ t_i\ t_j\} = \sum_{r=0}^{\infty} (\tfrac{1}{2})^r \frac{1}{r!}\ t_i^r\ t_j^r\ ,$$

it follows that

$$\phi(t_1,\ t_2,\ t_3,\ t_4) = \exp\{-\ \tfrac{1}{2}\ t_1^2\}\ \exp\{-\ \tfrac{1}{2}\ t_2^2\}\ \exp\{-\ \tfrac{1}{2}\ t_3^2\}\ \exp\{-\ \tfrac{1}{2}\ t_4^2\}$$

$$\cdot \sum_{r_1=0}^{\infty} \sum_{r_2=0}^{\infty} \sum_{r_3=0}^{\infty} (\tfrac{1}{2})^{r_1+r_2+r_3} \frac{1}{r_1!\ r_2!\ r_3!}\ t_1^{r_1}\ t_2^{r_1+r_2}\ t_3^{r_2+r_3}\ t_4^{r_3}\ .$$

By using the inversion formula for characteristic functions given by Crámer (1946), one can write

$$d = \int_{-\nu_1}^{\infty} \int_{-\nu_2}^{\infty} \int_{-\nu_3}^{\infty} \int_{-\nu_4}^{\infty} \text{MVN}(0,\ 0,\ 0,\ 0,\ \Sigma)\ dy_1\ dy_2\ dy_3\ dy_4$$

$$= \left(\frac{1}{2\pi}\right)^4 \int_{-\nu_1}^{\infty} dy_1 \int_{-\nu_2}^{\infty} dy_2 \int_{-\nu_3}^{\infty} dy_3 \int_{-\nu_4}^{\infty} dy_4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(t_1,\ t_2,\ t_3,\ t_4)$$

$$\cdot \exp\{-\ it_1 - it_2 - it_3 - it_4\}\ dt_1\ dt_2\ dt_3\ dt_4$$

$$= \left(\frac{1}{2\pi}\right)^4 \int_{-\nu_1}^{\infty} dy_1 \int_{-\nu_2}^{\infty} dy_2 \int_{-\nu_3}^{\infty} dy_3 \int_{-\nu_4}^{\infty} dy_4 \sum_{r_1=0}^{\infty} \sum_{r_2=0}^{\infty} \sum_{r_3=0}^{\infty} (\tfrac{1}{2})^{r_1+r_2+r_3}$$

$$\frac{1}{r_1!\ r_2!\ r_3!} \cdot \int_{-\infty}^{\infty} t_1^{r_1} \exp\{-\ \tfrac{1}{2}\ t_1^2 - it_1\ y_1\}\ dt_1$$

$$\cdot \int_{-\infty}^{\infty} t_2^{r_1+r_2} \exp\{-\ \tfrac{1}{2}\ t_2^2 - it_2\ y_2\}\ dt_2$$

$$\cdot \int_{-\infty}^{\infty} t_3^{r_2+r_3} \exp\{-\ \tfrac{1}{2}\ t_3^2 - it_3\ y_3\}\ dt_3$$

$$\cdot \int_{-\infty}^{\infty} t_4^{r_3} \exp\{-\ \tfrac{1}{2}\ t_4^2 - it_4\ y_4\}\ dt_4\ .$$

Now if we let $f(x)$ be the function $1/\sqrt{2\pi}\ e^{-\frac{1}{2}x^2}$ then $H_r(x)$, the r-th Hermite polynomial is defined as

$$H_r f(x) = \left(-\ \frac{d}{dx}\right)^r f(x)\ .$$

Also,

$$\int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2} t^2 - itx\}\, t^2\, dt = \frac{\partial^r}{\partial(-ix)^r} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2} t^2 - itx\}\, dt$$

$$= i^r \frac{\partial^2}{\partial x^2} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2} t^2 - itx\}\, dt$$

$$= i^r \frac{\partial^2}{\partial x^2} \sqrt{2\pi}\, e^{-\frac{1}{2} x i^2}$$

$$= (2\pi)(-i)^r H_r(x) f(x) ,$$

and

$$\int_{-\nu}^{\infty} dx\, H_r(x) f(x) = H_{r-1}(-\nu) f(-\nu) .$$

It follows directly that

$$d = \int_{-\nu_1}^{\infty} dy_1 \int_{-\nu_2}^{\infty} dy_2 \int_{-\nu_3}^{\infty} dy_3 \int_{-\nu_4}^{\infty} dy_4 \sum_{r_1=0}^{\infty} \sum_{r_2=0}^{\infty} \sum_{r_3=0}^{\infty} (\tfrac{1}{2})^{r_1+r_2+r_3}$$

$$\frac{1}{r_1!\, r_2!\, r_3!} \cdot (-i)^{r_1} H_r(y_1) f(y_1) \cdot (-i)^{r_1+r_2} H_{r_1+r_2}(y_2) f(y_2)$$

$$\cdot (-i)^{r_2+r_3} H_{r_2+r_3}(y_3) \cdot (-i)^{r_3} H_{r_3}(y_4) f(y_4)$$

$$= \sum_{r_1=0}^{\infty} \sum_{r_2=0}^{\infty} \sum_{r_3=0}^{\infty} (-\tfrac{1}{2})^{r_1+r_2+r_3} \frac{1}{r_1!\, r_2!\, r_3!} H_{r_1-1}(-\nu_1) f(-\nu_1)$$

$$\cdot H_{r_1+r_2-1}(-\nu_2) f(-\nu_2) H_{r_2+r_3-1}(-\nu_3) f(-\nu_3) H_{r_3-1}(-\nu_4) f(-\nu_4) .$$

The validity of the interchanging of the order of summation and integration in the above derivation follows from the fact that the series are uniformly convergent on any closed interval and the individual terms are continuous. It is shown by Kendall (1941) that the series obtained in the above manner does indeed converge, though in general the convergence depends on the size of the correlation and may be very slow. In the examples considered in this thesis, convergence was only moderately rapid and several hundred terms had to be computed. The necessary values for the expression $H_r(\nu)\ f(\nu)$ can be obtained from Tables of the Error Function and Its First Twenty Derivatives, published by the Harvard University Press (1952), though in this case they were computed using the relation

$H_r(\nu)\ f(\nu) = H_{r-1}(\nu)\ f(\nu) - (r-1)\ H_{r-2}(\nu)\ f(\nu)$, since such a large number of terms were needed. $H_0(\nu)\ f(\nu)$ is defined as $f(\nu)$ and $H_r(\nu)\ f(\nu)$ as $\int_\nu^\infty\ f(x)\ dx$. The accuracy of these computations was checked by evaluating a large number of terms in the area of interest and comparing the results with the tables. Agreement was to the sixth or seventh decimal place in almost all cases.

Only the case of 5 independent variables with unit variances has been discussed. The extension to more than 5 variables is obvious and need not be discussed.. Similarly the extension to correlated variables follows easily since the variance-covariance matrix of the vector of differences can be computed.

XIII. APPENDIX B

EVALUATION OF THE INCOMPLETE BETA FUNCTION

In the course of the computations for the power curves in Chapter VI it was necessary to evaluate the incomplete Beta function for a large number of parameter values. Rather than obtain these by interpolating in existing tables, it was decided to use a method outlined by Müller (1930) to compute the necessary values. In order to check the accuracy of the method the values were computed at a grid of points corresponding to points in existing tables and compared. The computed and the tabulated values agreed to five or more decimal places.

The formula use in the computations is obtained by truncating the series

$$I_x(u,v) = \frac{\Gamma(u+v)}{\Gamma(u+1)\,\Gamma(v)} x^u (1-x)^v \left\{ 1 + \frac{u+v}{u+1} x + \frac{(u+v)(u+v+1)}{(u+1)(u+2)} x^2 + \dots \right\} .$$

Since $0 \leq x \leq 1$ it follows that this series will always converge. The series will always converge. The infinite series was terminated at

$$\frac{(u+v)(u+v+1)\ \dots\ (u+v+38)}{(u+1)(u+2)\ \dots\ (u+39)} x^{39} .$$

The series expansion for $I_x(u,v)$ is derived by repeated use of the equation

$$\int x^{p-1} (1-x)^{q-1} dx = \frac{1}{p} x^p (1-x)^q + \frac{(p+q)}{p} \int x^p (1-x)^{q-1} dx .$$

This equation can be derived simply by writing

$$\int x^{p-1} (1-x)^{q-1} dx = \int x^{p-1} (1-x)^{q} dx + \int x^{p} (1-x)^{q-1} dx$$

and integrating by parts to obtain

$$\frac{1}{p} x^{p} (1-x)^{q} + \frac{q}{p} \int x^{p} (1-x)^{q-1} dx + \int x^{p} (1-x)^{q-1} dx$$

$$= \frac{1}{p} x^{p} (1-x)^{q} + \frac{q+p}{p} \int x^{p} (1-x)^{q-1} dx \ .$$

The necessary Gamma function values were built up in the computer, using the relation

$$\Gamma(x+1) = x\,\Gamma(x)$$

and reading in the initial values.